



**Breaking through the Threshold in Second Language Free Active
Vocabulary Acquisition**

by

Yinglai Yue, B.Arts., M.Arts.

Thesis Submitted in Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy

University of Tasmania

March 2018

Declaration of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief it contains no material previously published or written by another person except where due acknowledgement is made in the text or the thesis, nor does the thesis contain any material that infringes copyright.

March 2018

Statement regarding published work contained in thesis

The publishers of the papers comprising Chapter 2 hold the copyright for that content, and access to the material should be sought from the respective journals. The remaining non-published content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Bibliographic detail of the publication:

Yinglai, Y., & Si, F. (2016). Measurement of Vocabulary Knowledge: Problems and Solutions. In F. Si & F. W. Jill (Eds.), *What is Next in Educational Research?* (pp. 171-184). Netherlands: Sense Publishers.

March 2018

Statement of Co-authorship

The material in Chapter 2 has been published in a co-authored paper with Si Fan. The material was jointly conceptualised, the data analysis and a first draft were done by Yinglai Yue, and Si Fan edited the final draft.

Bibliographic detail of the publication:

Yinglai, Y., & Si, F. (2016). Measurement of Vocabulary Knowledge: Problems and Solutions. In F. Si & F. W. Jill (Eds.), *What is Next in Educational Research?* (pp. 171-184). Netherlands: Sense Publishers.

March 2018

Authority of Access

This thesis is not to be made available for loan or copying for two years following the date this statement was signed. Following that time the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

April 2018

Statement of Ethical Conduct

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator, and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

March 2018

Abstract

Free active vocabulary (FAV) refers to words that can be used automatically by people (Laufer & Paribakht, 1998). FAV is of great significance in second language (L2) study. However, FAV seems to be difficult to be acquired. Even advanced L2 learners' vocabulary use in speaking and writing tends to be limited. As early as in 1991, Laufer (1991) has found that there is a "threshold" in L2 free active vocabulary development. He found that during learners' study process, FAV fails to increase constantly with the growth of L2 proficiency. Observing this phenomenon, Laufer (1991) puts forward the "Threshold Hypothesis", proposing that L2 FAV would stagnate in L2 learning.

The aim of this research was twofold: first, to have a deeper understanding of FAV developmental process; secondly, if FAV experiences stagnation, to find possible reasons for the stagnation and possible ways to tackle the problem. To achieve this research aim, four sub-studies were conducted, which involved both quantitative and qualitative methods.

The first quantitative sub-study is a 12-month longitudinal research. Twenty-seven randomly selected third-year university students, whose major was English, were observed for 12 months. In the first and twelfth month, a composition of about 200 words was collected; meanwhile, tests for free active vocabulary, passive vocabulary (PV) and controlled active vocabulary (CAV) were conducted. Then all scores of the tests were put into the Statistical Package for the Social Sciences (SPSS) to be analysed. The purpose of this part of study was to obtain details on FAV development, and to explore the relationship between L2 learners' free active vocabulary knowledge and other types of vocabulary knowledge.

The second quantitative sub-study is a word association research exploring the cognitive mechanism in FAV development. First, a group of frequently produced words and seldom produced words were selected. Then word association tests were conducted to explore the lexical networks of the two kinds of vocabulary in the participants' mental lexicon. Their responses to the stimulus words were classified into different categories, and the number of responses in each category was counted. Then results were put into SPSS to be analysed. The purpose of the sub-study was to compare the lexical networks of the two types of vocabulary, which may shed some light on the lexical construction procedure in FAV development.

The first qualitative sub-study is the interview. It aimed to investigate whether students and teachers attached importance to FAV development, and whether effective teaching and learning strategies were adopted to promote FAV growth. Possible reasons for FAV stagnation were investigated as well.

The second qualitative sub-study is the documentary analysis. The course structure in the period of the 12-month longitudinal sub-study was examined. The English units taken during the period were analysed. The College English Syllabus for English Majors in China was examined as well. The purpose of the documentary analysis was to detect possible factors in course design that may affect FAV development.

Results of the sub-studies led to several findings. First, it was found that the FAV threshold phenomenon existed when L2 learners reached upper-intermediate to advanced L2 proficiency level. Three features were captured in the threshold phenomenon. The first feature is that the threshold phenomenon is long-lasting; the second feature is that the threshold phenomenon is "stubborn", as it does not grow

with the growth of PV and CAV; the third feature is that the FAV threshold tends to occur widely, as it occurs at most of the word frequency levels.

Secondly, it was found that both PV and CAV kept growing when L2 learners reached upper-intermediate to advanced proficiency level; FAV was not significantly correlated with PV and CAV in this stage, and the gap between PV and FAV, and between CAV and FAV became larger.

Thirdly, there are three tendencies in lexical network construction when a word is developing into FAV. The first tendency is that the word's lexical network tends to be semantised. The second tendency is that some types of nodes tend to increase in the process, while other types of nodes tend to decrease; and they change at different speeds. The third tendency is that the connections of the lexical network tend to get stronger and tighter. In summary, the lexical network tends to develop in a way that the word becomes easier to be accessed, activated, and retrieved.

Fourthly, it was found that there were possible factors for FAV stagnation in both teaching and learning. In teaching, the first factor is that teachers seem not to have given adequate attention to FAV in classroom instruction. The second factor is that little effort has been made to facilitate students in FAV development. The third factor is that there are inadequate units for output training in the curriculum. In learning, there are several possible factors for FAV stagnation as well. The first factor is that students' effort to improve FAV may not be enough, as the strategies they adopted were rather limited; sometimes the effort was not persistent, as it was given up later. The second factor is that there is insufficient intentional learning in FAV study.

Lastly, some effective teaching and learning strategies were detected from the research. The learning strategies include sentence making with the target word, note-

taking on the target word, putting down example sentences, collecting synonyms, etc. The effective teaching strategies include providing synonyms of the target word, negotiating word meanings, analysing the word's contexts, encouraging students to make sentences with the word, and raising students' attention to the word. In general, the research aim was fulfilled.

Acknowledgments

I would like to extend my sincere gratitude to the supervision team: Dr. Si Fan and Associate Professor Marion Myhill. Without your commitment, encouragement, consideration and support during my candidature, it would be impossible for me to complete this huge project. Especially I hope to thank Dr. Si Fan for her patience and guidance in the entire process. Thank you for taking a large amount of time and effort in my candidature. I would also like to thank the students and teaching staff who have contributed their insights and shared their research experiences with me. The support and assistance you gave me helped me go through this long journey smoothly.

To my loving parents and sister, my husband, parents-in-law and my son, thank you for giving me love and support that have accompanied me all the way along this journey. Especially I would like to thank my late father, thank you for giving me love, guidance, and courage for your whole life. I will love you forever.

Table of Contents

Chapter 1	Introduction	1
1.1	Research Motivation	1
1.2	Research Background	3
1.2.1	Threshold Phenomenon in Free Active Vocabulary Acquisition	3
1.2.2	Weaknesses of Past Research	3
1.3	Research Objectives and Research Questions	5
1.3.1	Research Objective One and Corresponding Research Questions	5
1.3.2	Research Objective Two and Corresponding Research Questions.....	6
1.3.3	Research Objective Three and Corresponding Research Questions.....	6
1.3.4	Research Objective Four and Corresponding Research Questions.....	6
1.3.5	Research Objective Five and Corresponding Research Questions	7
1.4	Participants.....	7
1.5	Research Methods.....	9
1.6	Ethical Considerations	10
1.7	Research Significance	11
1.7.1	Theoretical Significance	11
1.7.2	Practical Significance.....	12
1.8	Outline of the Thesis.....	12
Chapter 2	Literature Review	16
2.1	Introduction.....	16
2.2	The Classification of Vocabulary Knowledge.....	17
2.2.1	Classifying Vocabulary Knowledge from Degree Perspective	17
2.2.2	Classifying Vocabulary Knowledge from Dimension Perspective	21

2.2.3	Classifying Vocabulary Knowledge from both Degree and Dimension	
	Perspective	22
2.3	The Measurement of Vocabulary Knowledge	24
2.3.1	Breadth and Depth of Vocabulary Knowledge	24
2.3.2	Measuring Vocabulary Knowledge of Individual Words	24
2.3.2.1	Measuring the Dimensional Knowledge of Individual Words	24
2.3.2.2	Measuring the Depth of Knowledge for Individual Words	27
2.3.3	Measuring Vocabulary Knowledge of the Entire Vocabulary	29
2.3.3.1	Measuring the Breadth of Passive Vocabulary	29
2.3.3.2	Measuring the Breadth of Controlled Active Vocabulary	31
2.3.3.3	Measuring the Breadth of Free Active Vocabulary	32
2.4	Mental Lexicon	37
2.4.1	Definition of Mental Lexicon	37
2.4.2	Nature of L2 Mental Lexicon	38
2.4.3	The Organisational Structure of Mental Lexicon	43
2.4.3.1	The Hierarchy Model	44
2.4.3.2	The Spreading-activation Model	46
2.4.4	Bilinguals' Lexical Retrieval	50
2.4.5	Significance of Mental Lexicon Study to this Research	51
2.4.6	Word Association Tests in Mental Lexicon Study	52
2.4.6.1	The Fundamental Assumption behind WA Tests	53
2.4.6.2	Selection of Stimulus Words in WA Tests	53
2.4.6.3	The Categorisation of Responses	60
2.4.6.4	The Response Categorisation in this Research	69
2.4.6.5	Weaknesses of the New Categorisation	72

2.5	The Threshold Phenomenon in Free Active Vocabulary Acquisition	73
2.6	The Working Definitions and Terminology in the Research.....	76
2.7	Conclusion	77
Chapter 3	Methodology.....	78
3.1	Introduction.....	78
3.2	Research Aim and Objectives	78
3.3	Research Approach and Methods	79
3.4	Instruments and Data Collection.....	83
3.4.1	Instruments and Data Collection in the Longitudinal Sub-study.....	83
3.4.2	Instruments and Data Collection in the Word Association Sub-study	86
3.4.3	Instruments and Data Collection in the Interview	86
3.4.4	Instruments and Data Collection in the Documentary Analysis	89
3.5	Pilot Study.....	90
3.6	Sampling and Participants.....	91
3.6.1	Sampling and Participants in the Longitudinal Sub-study.....	91
3.6.2	Sampling and Participants in the Word Association Sub-study	95
3.6.3	Sampling and Participants of the Interview	96
3.7	Data Analysis	97
3.7.1	Quantitative Data Analysis	98
3.7.2	Qualitative Data Analysis	98
3.8	Conclusion	99
Chapter 4	The Developmental Process of Free Active Vocabulary	101
4.1	Introduction.....	101
4.2	Research Questions	101
4.3	Research Method	102

4.3.1	Instruments.....	102
4.3.1.1	Test of Receptive Vocabulary.....	102
4.3.1.2	Test of Controlled Active Vocabulary.....	106
4.3.1.3	Test of Free Active Vocabulary.....	107
4.3.1.3.1	Lex30 Test.....	107
4.3.1.3.2	Writing Tasks.....	108
4.3.2	Data Collection Procedures.....	113
4.3.2.1	Pilot Study.....	113
4.3.2.2	Formal Data Collection.....	114
4.3.3	Data Preparation.....	116
4.3.3.1	Selection of Valid Data.....	116
4.3.3.2	Scoring of PVLТ.....	116
4.3.3.3	Scoring of CAVLT.....	116
4.3.3.4	Scoring of Lex30 Test.....	117
4.3.3.5	Scoring of Writing Tasks.....	119
4.3.4	Data Analysis.....	122
4.4	Results.....	123
4.4.1	The Developmental Process of FAV in the 12-month Period of Time.....	123
4.4.2	The Correlation of FAV with Passive Vocabulary and Controlled Active Vocabulary.....	126
4.5	Discussion.....	130
4.5.1	The Developmental Process of FAV.....	130
4.5.2	The Relationship of FAV with Passive Vocabulary and Controlled Active Vocabulary.....	132
4.6	Conclusion.....	136

Chapter 5 Comparison of Lexical Network between Frequently and Seldom

Produced Words.....	137
5.1 Introduction.....	137
5.2 Research Questions	138
5.3 Research Method	139
5.3.1 Instruments.....	139
5.3.1.1 Vocabulary Knowledge Scale Test.....	139
5.3.1.2 Word Association Test.....	140
5.3.2 Data Collection Procedures.....	144
5.3.2.1 Pilot Study.....	144
5.3.2.2 Formal Data Collection.....	145
5.3.3 Data Preparation.....	147
5.3.3.1 Selection of Valid Data	147
5.3.3.2 Scoring the VKS Tests.....	148
5.3.3.3 Scoring the Word Association Tests.....	148
5.3.4 Data Analysis	148
5.4 Results.....	149
5.4.1 Semantic and Non-semantic Distribution of SPW and FPW.....	150
5.4.2 Distribution of Different Types of Semantic Nodes in SPW and FPW.....	152
5.4.3 Distribution of Different Types of Non-Semantic Nodes in SPW and FPW	159
5.4.4 Qualitative Data Analysis	162
5.5 Discussion	163
5.5.1 Differences of Lexical Representation between SPW and FPW	163
5.5.2 The Lexical Construction Procedure in SPW-to-FPW Process.....	172

5.5.3	Pedagogical Implications	175
5.6	Conclusion	177
Chapter 6	Interview	179
6.1	Introduction.....	179
6.2	Aspects of Interview Questions	179
6.3	Research Method	180
6.3.1	Instrument	180
6.3.2	Data Collection Procedure	181
6.3.2.1	Pilot Study.....	181
6.3.2.2	Formal Data Collection.....	182
6.3.3	Data Preparation.....	182
6.4	Results and Discussion	182
6.5	Conclusion	193
Chapter 7	Documentary Analysis	197
7.1	Introduction.....	197
7.2	Research Questions	197
7.3	Research Method	198
7.3.1	Data Collection	198
7.3.2	Data Preparation.....	198
7.4	Results and Discussion	198
7.5	Conclusion	207
Chapter 8	Discussions, Findings, Implications and Recommendations	209
8.1	Introduction.....	209
8.2	Discussions	210
8.3	Major Findings.....	216

8.3.1	Findings to Research Objective One: the Developmental Process of FAV.....	216
8.3.2	Findings to Research Objective Two: the Relationship between FAV and Other Types of Vocabulary Knowledge	217
8.3.3	Findings to Research Objective Three: the Lexical Network Construction in FAV Development	219
8.3.4	Findings to Research Objective Four: Possible Factors Related to FAV Stagnation	222
8.3.5	Findings to Research Objective Five: Effective Strategies to Tackle the Threshold Phenomenon	224
8.4	Implications.....	225
8.4.1	Theoretical Implications	225
8.4.2	Pedagogical Implications	226
8.5	Recommendations.....	228
8.5.1	Tactics to Strengthen Network Building.....	228
8.5.2	Tactics to Strengthen Intentional FAV Learning.....	231
8.6	Conclusion	233
Chapter 9	Conclusion.....	234
9.1	Introduction.....	234
9.2	The Research Journey	234
9.3	Summary of the Major Findings	237
9.4	Significance of the Research.....	239
9.5	Possible Directions for Future Research.....	240
9.6	Conclusion	243
	References.....	245

Appendix A The Passive Vocabulary Levels Test	269
Appendix B The Controlled Active Vocabulary Levels Test	274
Appendix C The Lex30 Test.....	278
Appendix D Writing Task One	280
Appendix E Writing Task Two.....	281
Appendix F Word Association Test One	282
Appendix G Word Association Test Two.....	283

List of Tables

Table 2.1. Dimensions in the Vocabulary Knowledge Framework by Richards (1976)	22
Table 2.2. The Vocabulary Knowledge Classification by Webb (2005, 2007, 2009)	23
Table 2.3. The Vocabulary Knowledge Measurement Instruments by Webb (2007, p. 55) ...	25
Table 2.4. An Example of Vocabulary Knowledge Measurement by Gina et al. (2011, p.52)	27
Table 2.5. The Concreteness Scale for Words of Different Lexical Classes (P. Zhang, 2010a)	59
Table 3.1. Key Information on the Quantitative Sub-studies	81
Table 3.2. Key Information on the Qualitative Sub-studies	82
Table 4.2. Comparison of Lex30 and LFP in Terms of Word Types between the First Month and the Twelfth Month.....	123
Table 4.3. Comparison of LFP in Terms of Word Families between the First Month and the Twelfth Month	124
Table 4.4. Comparison of PVLTL Scores between the First Month and the Twelfth Month..	126
Table 4.5. Comparison of CAVLT Scores between the First and the Twelfth Month	127
Table 5.1. Key Information on Selected FPW	144
Table 5.2. Key Information on Selected SPW	144
Table 5.3. Comparison of Semantic and Non-semantic Responses in FPW and SPW	150
Table 5.4. Comparison of Semantic and Non-semantic Responses between FPW and SPW	151
Table 5.5. Descriptive Statistics for Different Types of Semantic Responses in FPW and SPW	152
Table 5.6. Comparison of Different Types of Semantic Responses between FPW and SPW	153

Table 5.7. Descriptive Statistics for Different Types of Paradigmatic Responses in FPW and SPW	155
Table 5.8. Comparison of Different Types of Paradigmatic Responses between FPW and SPW	156
Table 5.9. Comparison of Different Types of Syntagmatic Responses in FPW and SPW....	158
Table 5.10. Comparison of Different Types of Syntagmatic Responses between FPW and SPW	159
Table 5.11. Comparison of Different Types of Non-semantic Responses between FPW and SPW	160
Table 5.12. Descriptive Statistics for Different Types of Non-semantic Responses in FPW and SPW.....	161
Table 6.1. Aspects and Corresponding Questions in the Interview	180
Table 7.1. Discipline-related Units	198

List of Figures

Figure 2.1. The Hierarchy Model by Collins and Quillian (1969, p.241)	44
Figure 2.2. The Spreading-activation Model by Collins and Loftus (1975, p. 412)	46
Figure 2.3. The Modified Spreading-activation Model by Bock and Levelt (1994, p. 951) ..	49
Figure 2.4. The categorisation of responses in WA tests.....	70
Figure 3.1. The research approach and methods adopted in this study	80

Chapter 1

Introduction

1.1 Research Motivation

Free active vocabulary (FAV) refers to words that can be used automatically by people (Laufer & Paribakht, 1998). FAV is of great significance in second language (L2) study. To some extent, it determines the quality of speaking and writing (Astika, 1993; Coady, Magoto, Hubbard, Graney, & Mokhtari, 1993; Engber, 1995; Laufer, 1994; Yamamoto, 2011). However, it seems that FAV is difficult to be acquired, as even advanced English as a second language (ESL) learners' vocabulary use in speaking and writing tends to be limited. The weakness in FAV usage is reflected in the teaching experience of the researcher. In her eight-year ESL teaching career, the researcher found that her English-major students, including those senior students who had been studying English as a second language for over ten years, were confined to the most common, unnuanced words in writing. For example, when students needed to express "laugh repeatedly in a quiet but uncontrollable way", few of them could think of the word "giggle"; instead, most of them would use the word "laugh" or "smile", even if they knew that in this way the nuanced meaning would be lost. For another example, when they wanted to say that a tiger is stepping impatiently in cage,

few of them could think of the verb “pace”. A more common, basic word “walk” was usually used as a substitute, even if they knew “walk” may not be the most accurate and appropriate word in this case.

This researcher’s personal experience is consistent with the finding of Jullian (2000) that although upper-intermediate to advanced L2 learners can communicate with correct grammar, they usually overuse core words. The overuse of general core words in speaking and writing not only results in a poor and child-like discourse, but also makes them unable to describe different moods and connotations with specific semantic loads (Jullian, 2000). Also, L2 learners’ experience reflects this problem as well. Yu (2011) conducted a survey among 88 English-major students who were in the second year of university study. The survey shows that 87.3% of them felt that their FAV was too limited, which made them unable to express what they wanted to express (Yu, 2011).

In addition to personal experiences, some empirical studies also show that L2 learners’ FAV is far from adequate, as there is a big gap between native speakers’ (NSs) and L2 learners’ FAV size. Cobb (2003) found that in NSs’ expository texts, 70% of the words fell into the first 1,000 frequency zone; in NSs’ oral conversational texts, 80% of the words fell into this zone. By contrast, as high as 90% of words used by advanced ESL learners in expository and argumentative writings were from the first 1,000 frequency zone. Cobb (2003) therefore proposed the Overuse Hypothesis that even advanced L2 learners tend to overuse general, unnuanced words. Laufer (1994) found there is a big gap between NSs and L2 learners in FAV usage as well. Laufer (1994) compared words used by a group of 18-year-old NSs, who had not entered college yet, and a group of first-year English-major university students in Israel. He found that in NSs’ argumentative writing on general topics, 25% to 28%

lexical items fell into the beyond-2,000 word frequency range; by contrast, the percentage of Israeli students was only half of the NSs. The salient gap between NSs and L2 learners of various proficiency levels in FAV usage demonstrates that L2 learners' FAV size is far from being adequate. The limitedness of FAV becomes a factor that hinders the improvement of L2 learners' writing and speaking quality.

1.2 Research Background

1.2.1 Threshold Phenomenon in Free Active Vocabulary Acquisition

As early as in 1991, Laufer (1991) has found that there is a "threshold" in L2 FAV development. He found that during learners' study process, FAV fails to increase constantly with the growth of L2 proficiency; its growth stops after developing to a certain level. Observing this phenomenon, Laufer (1991) puts forward the "Threshold Hypothesis", proposing that L2 FAV would stagnate in L2 learning. After Laufer's research, a great deal of studies (e.g. Gu & Li, 2013; Huang, 2012; Laufer, 1994; Lu, 2008; Tan, 2006; X. Wu & Chen, 2000; G. Zhang, Han, & Zhu, 2005; Y. Zhao, 2011) proved the existence of the "threshold", or as Cobb (2003, p. 403) called it, the "ceiling effect". Meanwhile, as is stated in Section 1.1, there is a big gap between NSs' and L2 learners' FAV size. If the "threshold" problem cannot be solved, the efficiency of L2 learners' language study will be seriously affected.

1.2.2 Weaknesses of Past Research

Although the threshold phenomenon in L2 FAV acquisition has drawn increasing attention, relevant research is far from thorough. The main weaknesses include: Firstly, numerous studies have proved the existence of "threshold", yet most of them

did not pay specific attention to the details of the threshold. For instance, when the phenomenon occurs, does it occur at all frequency levels? Or does it occur at some of the frequency levels? What is the Lexical Frequency Profile (LFP) like when the threshold occurs?

Secondly, previous studies have tried to detect reasons for the emergence of threshold, but conclusions are usually unilateral. Ineffective teaching and learning strategies are always blamed for the failure of FAV's constant growth (Cui & Wang, 2006; Gu & Li, 2013; Lu, 2008; Tan, 2006; G. Zhang et al., 2005); lack of motivation is blamed as well (Yang, 2007). It seems that cognitive causes have seldom been investigated. In addition, those factors being blamed are only speculations, not backed up by empirical studies. Little attention has been paid to learners' cognitive mechanism, which may help uncover the root causes for FAV stagnation. Meara (1990) pointed out that exploring the cognitive process in learners' vocabulary acquisition is an important part in L2 vocabulary acquisition study. The fact that, L2 learners may recognise a word well but may never be able to automatically use that word, indicates that the cognitive process in vocabulary acquisition may be more complicated than what people thought. However, how lexical networks affect words' retrieval and activation is still, to a great extent, elusive to people.

Thirdly, relevant studies are primarily cross-sectional (e.g. Cui & Wang, 2006; Gu & Li, 2013; Huang, 2012; Laufer, 1998; Liu, 2001; Lu, 2008; Tan, 2006, 2007; Wen, 2006; G. Zhang et al., 2005). Due to the large quantity of time and energy needed in a longitudinal study, most relevant research adopted cross-sectional study instead of longitudinal study to explore the developmental pattern of FAV. In addition, few studies have adopted long-time longitudinal study and track the development of different types of vocabulary knowledge simultaneously. Such studies are needed as

they can help researchers obtain more accurate information on FAV development.

Fourthly, literature review shows that almost no research has investigated the courses students took and checked whether some factors in course design may affect FAV progress. As English is a foreign language in China, class instruction is one of the major sources from which students obtain English training. Therefore, the course design is a significant part in L2 learning and should not be overlooked in relevant research.

The present research attempts to overcome the four weaknesses stated above. A 12-month longitudinal sub-study was conducted to track L2 learners' free active vocabulary development. Meanwhile, the development of passive vocabulary (PV) and controlled active vocabulary (CAV) was tracked as well, so that the relationships between FAV and PV and between FAV and CAV could be investigated. The lexical network and the units participants took were also examined.

1.3 Research Objectives and Research Questions

1.3.1 Research Objective One and Corresponding Research Questions

The first research objective of this study is to investigate the developmental process of FAV when L2 learners reach the upper-intermediate to advanced L2 proficiency level.

The following questions are asked in relation to the research objective:

- What is the development pattern of FAV when L2 learners reach upper-intermediate to advanced proficiency level?
- If the threshold phenomenon occurs in the FAV development, what are the features of the threshold phenomenon?

1.3.2 Research Objective Two and Corresponding Research Questions

The second research objective is to investigate the relationship between FAV and PV, and the relationship between FAV and CAV. There are four research questions associated with this research objective:

- What is the developmental process of PV when L2 learners reach upper-intermediate to advanced L2 proficiency level?
- What is the relationship between FAV and PV?
- What is the developmental process of CAV when L2 learners reach upper-intermediate to advanced L2 proficiency level?
- What is the relationship between FAV and CAV?

1.3.3 Research Objective Three and Corresponding Research Questions

The third research objective is to explore how lexical network is constructed in the process of FAV development. Based on the research objective, one research question is proposed:

- What are the differences (if any) of the mental lexicon organisation between frequently produced words and seldom produced words?

1.3.4 Research Objective Four and Corresponding Research Questions

The fourth research objective is to find possible factors in learning and teaching that may affect FAV development. Based on the research objective, two questions are designed:

- What are the possible factors in teaching that affect FAV development?

- What are the possible factors in learning that affect FAV development?

1.3.5 Research Objective Five and Corresponding Research Questions

If the threshold phenomenon is detected in the longitudinal sub-study, a fifth research objective and corresponding research questions will be proposed. The fifth research objective is to find effective ways to tackle the FAV threshold phenomenon. The research question is:

- What are the effective strategies (if any) to tackle FAV threshold phenomenon?

1.4 Participants

The participants of the study were English-major students at a public university in China. All the participants experienced standard education in China (primary school, junior high, senior high, and university). These students were at the end of Year Three when the research was conducted. With English being their major, they were considered to have a high level of motivation to study the language. The students' ages were mostly 21 to 22. It was believed that the subjects were suitable for this study, since being English-major students in the third year at the university, they had been studying English for at least nine years (six years in middle school before entering college), and had done a number of discipline-related units. The courses that the participants took were designed according to The Syllabus of English Courses for College English Majors by the Chinese Ministry of Education. The syllabus requires that in the four years' study, English-major students should receive 2,000 to 2,200 academic hours' discipline-related instruction. Also, according to the syllabus, by the

end of the third-year study, the English-major students should reach the Sixth Proficiency Level. The syllabus provides a detailed description on the Six Proficiency Level. In speaking, learners should be able to pronounce English words correctly, and they should be able to speak English naturally and fluently; they should be able to introduce China's well-known and historic sceneries fluently and correctly, and introduce China's policies and conditions fluently as well; they should be able to express their views systematically, coherently, and in an in-depth way. In vocabulary size, they should be able to recognise 7,000 to 9,000 English words. In listening, they should be able to understand the news report from the channels of English-speaking countries; they should be able to have a dictation of recordings that speak 150 words per minute, and the error rate should be lower than 6%. In reading, students should be able to read the politics articles as difficult as articles in magazines of Times and New York Time; they should be able to read the original English literature as difficult as *the Great Gatsby*, and they should be able to read the biographies as difficult as *the Rise and Fall of the Third Reich*. In addition to understanding the works, they should be able to analyse the views of the author, the structure, the genre and styles of the works. The reading speed should reach 140 to 180 words per minute, and the understanding rate should be no less than 75%. Within 5 minutes, they should be able to read an article of 1,300 words, and after reading they should be able to tell the gist of the article. In writing, students are required to be able to write summaries of stories, book reports, course papers, and formal letters. The language they use in writing should be correct and appropriate, and the writing should reflect in-depth thinking. According to the syllabus, the writing speed should be 250 to 300 words per minute. Meanwhile, students are required to be familiar with the use of all sorts of encyclopedia, including the Encyclopedia Britannica and the Encyclopedia

Americana. The syllabus requires that students should be able to independently seek answers for questions in language and in other world knowledge. At the end of the fourth-year study, the proficiency level required by the syllabus is even higher. Based on the requirement of the syllabus, it is postulated that the participants in this research are representative of upper-intermediate to advanced ESL students in China.

1.5 Research Methods

The study was conducted using a mixed approach, involving both quantitative and qualitative methods. The quantitative sub-studies include:

- (1) A 12-month longitudinal sub-study. Twenty to 40 randomly selected third-year university students, whose major was English, were observed for 12 months. In the first and twelfth month, a composition of about 200 words was collected; meanwhile, tests for FAV, PA and CAV were conducted. Then all scores of the tests were put into SPSS and analysed. The purpose of this part of study was to obtain details on FAV development, and to explore the relationship between L2 learners' FAV knowledge and other vocabulary knowledge.
- (2) A sub-study on the cognitive mechanism in FAV development. First, a group of frequently produced words and seldom produced words were selected. The selection process will be introduced in Chapter 5. Then word association tests were conducted to explore the lexical networks of the two kinds of vocabulary in subjects' mental lexicon. Responses to the stimulus words were classified into different categories, and the number of responses in each category was counted. Then results were put into SPSS and analysed. The purpose of the sub-study was to compare the lexical networks of the two types of vocabulary,

which may shed some light on the lexical construction procedure in FAV development.

The qualitative research includes:

- (1) Interviews were conducted. The interview aimed to investigate whether students and teachers attached importance to FAV development, and whether effective teaching and learning strategies were adopted to promote FAV growth. Possible reasons for FAV stagnation were investigated as well.
- (2) A documentary analysis was conducted. The course structure in the period of the 12-month longitudinal sub-study was examined. The English units taken in the period were analysed. The Syllabus of English Courses for College English Majors by the Chinese Ministry of Education (2000) in China was examined as well. The purpose of the documentary analysis was to detect possible factors in course design that may affect FAV development.

1.6 Ethical Considerations

The researcher of this study held ethical awareness throughout the process of data collection, data analysis, and data interpretation. The study was given full ethical approval by the ethics committee of the University of Tasmania on March, 10th, 2015 (Ethics Ref No: H0014755).

The research is of minimal risk from the ethics point of view. The research was conducted in the student investigator's home country of China, which held no potential risk for the researcher to conduct the research. In addition, the student investigator had an advantage of being a Chinese citizen with the knowledge of Chinese social, cultural and religious values, mannerisms, and most importantly, the official language Mandarin.

There was no risk foreseen for the participants of this research. The factors examined were related specifically to their vocabulary study, which did not pose any risk to the participants' professional or social life. The participants were individually interviewed, so that their confidentiality could be protected. In addition, the recordings of the interviews were coded to be Interviewee A, Interviewee B, and Interviewee C, so that others could not connect the recordings with the participants. No question related to their culture or religion was asked. In addition, the tests conducted in the study were English vocabulary knowledge tests and word association tests, which had no sensitive or private issues.

No data was collected or used without the participants' consent. Prior to the tests and interviews, information sheets and consent forms were handed out to participants. They were asked to read the information sheets carefully and then sign the consent forms if they agreed to do the interview or tests. The research involved no deception of participants. The purposes of the research were not concealed; no observation or audio recording was done without consent. Moreover, the participants were made aware that they had the right not to answer the questions or stop the data collection procedure at any point in the process. Data were collected in university settings so that there were no safety issues.

1.7 Research Significance

1.7.1 Theoretical Significance

The study further testifies and develops the "Threshold Hypothesis" put forward by Laufer (1991). In addition to testifying the hypothesis, more detailed information on the FAV threshold phenomenon was obtained. The study also helps researchers have a

better understanding of FAV stagnation by exploring the different lexical organisation of frequently produced words and seldom produced words. This may shed some light on the cognitive mechanism of FAV development.

1.7.2 Practical Significance

The study helps researchers and ESL teachers obtain more knowledge on FAV development and a better understanding of the FAV threshold phenomenon. Moreover, effective teaching and learning strategies to tackle FAV stagnation are detected. Teachers and learners can obtain more guidance from this research to tackle the problem.

1.8 Outline of the Thesis

The thesis, which is a report of the entire research project, contains nine chapters in total. The nine chapters are: Introduction, Literature Review, Methodology, the Developmental Process of Free Active Vocabulary, Comparison of Lexical Network between Frequently Produced Words and Seldom Produced Words, Interview, Documentary Analysis, Findings, Implications and Recommendations, and Conclusion. A summary of the content of each chapter is presented below.

Chapter 1: Introduction

This chapter serves as an introduction to the entire research project. To justify the necessity to do this research, the chapter first presents research motivation, findings in past research, and limitations of past research. Then it describes the research objectives of this study. The research questions are specified as well, followed by a brief description of participants and research methods. The chapter then presents ethical considerations. The significance of the research, which includes both

theoretical and practical significance, is discussed. The chapter ends with the outline of the entire thesis.

Chapter 2: Literature Review

This chapter reviews a range of literature related to the research topic. The aim of the literature review is to provide the project with theoretical foundations. It firstly examines and discusses the classifications of vocabulary knowledge. This step leads to definitions of the key concepts used in this research, including passive vocabulary, controlled active vocabulary, and free active vocabulary. The chapter then examines instruments measuring different types of vocabulary knowledge, which helps to select suitable instruments for this research. Relevant literature on mental lexicon is reviewed as well, which provides theoretical background for the lexical network sub-study in this research. Moreover, the chapter has an in-depth review of past studies on L2 FAV threshold phenomenon; findings and limitations of those studies are investigated and discussed.

Chapter 3: Methodology

This chapter is an overview of the methodology adopted in the research. It first introduces the research aim and objectives, then describes the mixed-method approach adopted in the study. The research instruments used in the data collection, together with participants and sampling process are described as well. The pilot studies and the data analysis methods are also generally introduced. To present in a more cohesive way, the details on the pilot study for each sub-study and data analysis methods used in each sub-study are provided in the following chapters with the formal sub-study.

Chapter 4: the Developmental Process of Free Active Vocabulary

This chapter tracks the developmental process of FAV in a 12-month period of time. It

also investigates the relationship between FAV and other types of vocabulary, i.e., the passive vocabulary and controlled active vocabulary. The chapter first poses research questions, then presents research method, including instruments, data collection, data preparation and data analysis. This is followed by a discussion of the results. Conclusions are made at the end of the chapter.

Chapter 5: Comparison of Lexical Network between Frequently Produced Words and Seldom Produced Words

This chapter investigates and compares the lexical networks of frequently produced words and seldom produced words. The differences between the two (if any) may shed some light on the lexical network construction in FAV development. Moreover, detecting the differences may help researchers and ESL teachers find effective ways to tackle the stagnation problem. This chapter first raises research questions, then presents research methods, including research instruments, data collection procedures, and data analysis. At the end of the chapter, results are discussed, and conclusions are made.

Chapter 6: Interview

This chapter aims to investigate whether students and teachers attach importance to FAV development, and whether effective teaching and learning strategies are adopted to promote FAV growth. Possible reasons for FAV stagnation will be investigated as well. The chapter first poses research aims, then presents research method, including instruments, data collection, data preparation and data analysis. At the end of the chapter conclusions on the findings of this sub-study are made.

Chapter 7: Documentary Analysis

This chapter investigates the course structure students took in the period of the 12-month longitudinal sub-study. The Syllabus of English Courses for College English

Majors in China by the Chinese Ministry of Education (2000) was examined as well. The aim of the documentary analysis is to examine whether there are factors in course design that may affect students' FAV development. The chapter first poses research questions, then presents research method, including data collection, data preparation and data analysis. At the end of the chapter conclusions on the findings of this sub-study are made.

Chapter 8: Discussions, Findings, Implications and Recommendations

In light of the previous chapters, this chapter examines whether the five research objectives have been addressed and achieved through the research. Discussions are also made to examine whether the research questions have been answered satisfactorily. It first has a general discussion on the results, then presents findings pertaining to the research objectives and research questions, then discusses the theoretical and pedagogical implications of the findings. Lastly, recommendations for FAV development are provided.

Chapter 9: Conclusion

This chapter is a conclusion of the thesis. It first provides an overview of the entire research, which involves a reflection of the research journey and achievements from the journey. It also summarises the major findings and discusses the significances of the research. Then limitations of the research are disclosed, and possible directions for future study are provided.

Chapter 2 has been
removed for copyright or
proprietary reasons.

It has been published as: Yinglai, Y., Si, F. (2016). Measurement of vocabulary knowledge: problems and solutions. In F. Si, F. W. Jill (Eds.), What is next in educational research? (pp. 171-184). Netherlands: Sense Publishers

Chapter 3

Methodology

3.1 Introduction

This chapter is an overview of the methodology adopted in the research. It first introduces the research aim and objectives, and then describes the mixed-method approach adopted in the study. The research instruments used in the data collection, together with the participants and sampling process, are described as well. The pilot study and the data analysis methods are also introduced.

3.2 Research Aim and Objectives

The aim of the research was twofold: firstly, to have a deeper understanding of the free active vocabulary (FAV) developmental process; and secondly, if FAV experiences stagnation, to find possible reasons for the stagnation and possible ways to tackle the problem. The research aim served as an axis around which the entire research project was designed and carried out. Following the method by Leedy and Ormrod (2005), the research aim was broken into sub-problems to make it more manageable. The sub-problems are stated as research objectives in this study, which are explained below.

Research Objective 1: To investigate the developmental process of FAV in upper-intermediate to advanced L2 learners. This objective was to track the development of FAV in L2 learning, to testify the FAV threshold hypothesis proposed by Laufer (1998), and to explore the features of the threshold phenomenon if it exists.

Research Objective 2: To investigate the relationship between FAV and passive vocabulary (PV), and the relationship between FAV and controlled active vocabulary (CAV). This research objective was to compare the development pattern of FAV with that of PV and CAV, and to investigate the correlation between FAV and the other two types of vocabulary knowledge.

Research Objective 3: To explore how lexical network in mental lexicon is constructed in the process of FAV development. This research objective was to detect possible factors in mental lexicon construction that help promote FAV development.

Research Objective 4: To identify factors in learning and teaching that may affect FAV development. This research objective was to investigate whether some teaching and learning activities, and factors in course design, may affect FAV acquisition.

Research Objective 5: If the threshold phenomenon is detected in the longitudinal sub-study, a fifth research objective and corresponding research questions will be proposed. The fifth research objective is to find effective ways to tackle the FAV threshold phenomenon.

3.3 Research Approach and Methods

This research adopted a mixed-method approach, which involved both quantitative and qualitative methods. Following Tashakkori and Teddlie (1998), the quantitative and qualitative methods were adopted separately in different research phases. Compared with the approach that only uses either quantitative or qualitative method,

the advantage of mixed-method approach is that it has complementary strengths and non-overlapping weaknesses, which helps improve the research quality (Johnson & Christensen, 2014). In this research the quantitative sub-studies included a longitudinal sub-study and a word association sub-study. The qualitative sub-studies included interviews and documentary analysis. It was believed that the mix-method approach enabled the researcher to obtain more significant ideas and deeper investigation into the issue. The research approach and methods are demonstrated in Figure 3.1.

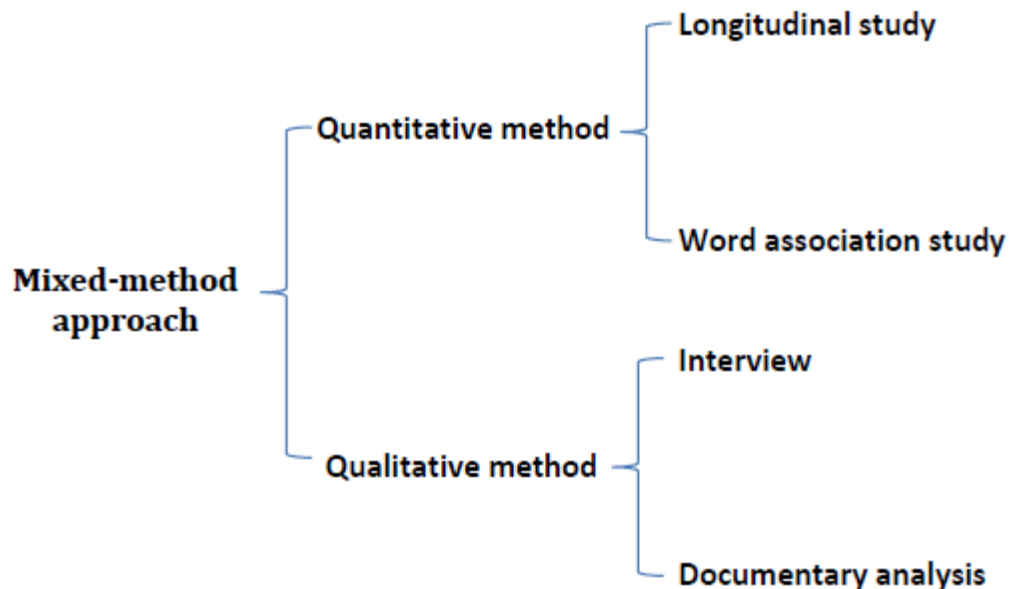


Figure 3.1. The research approach and methods adopted in this study

As demonstrated in Figure 3.1, there were two sub-studies adopting quantitative method, and two sub-studies adopting qualitative method in this research. The key information of the sub-studies is demonstrated in Table 3.1 and Table 3.2.

Table 3.1. *Key Information on the Quantitative Sub-studies*

Sub-studies	Participants	Activity and instruments	Time	Purpose	Research questions
The longitudinal sub-study	27 English-major undergraduate students	To conduct a battery of vocabulary knowledge tests, including Passive Vocabulary Levels Test, Controlled Active Vocabulary Test, Lex30 test, and writing tasks.	May, 2015	(1)To track the development of FAV in L2 learning, to testify the FAV threshold hypothesis proposed by Laufer (1998), and to explore the features of the threshold phenomenon if it exists ;	(1)What is the development pattern of FAV when L2 learners reach upper-intermediate to advanced proficiency level?
		To conduct a battery of vocabulary knowledge tests, including Passive Vocabulary Levels Test, Controlled Active Vocabulary Test, Lex30 test, and writing tasks.	May, 2016	(2) To compare the development pattern of FAV with that of PV and CAV, and to investigate the correlation between FAV and the other two types of vocabulary knowledge.	(2)If the threshold phenomenon occurs in the FAV development, what are the features of the threshold phenomenon? (3)What is the developmental process of PV when L2 learners reach upper-intermediate to advanced L2 proficiency level? (4)What is the relationship between FAV and PV? (5)What is the developmental process of CAV when L2 learners reach upper-intermediate to advanced L2 proficiency level?

					(6)What is the relationship between FAV and CAV?
The word association sub-study	30 English-major undergraduate students	(1)To do vocabulary knowledge scale test 1; (2)To do word association test 1	May, 2015	To detect possible factors in mental lexicon construction that help promote FAV development.	(1)What are the differences (if any) of the mental lexicon organisation between frequently produced words and seldom produced words?
		(1)To do vocabulary knowledge scale test 2; (2)To do word association test 2	June, 2015		

Table 3.2. *Key Information on the Qualitative Sub-studies*

Sub-studies	Participants	Activity	Time	Purpose	Research questions
Interview	Three English-major undergraduate students	To do an interview with the student investigator	June, 2016	To investigate whether some teaching and learning activities may affect FAV acquisition.	(1)What are the possible factors in teaching that may affect FAV development? (2)What are the possible factors in learning that may affect FAV development? (3) What are the effective strategies (if any) to tackle FAV threshold phenomenon?

Documentary analysis		To collect and analyse the course structure in the first semester of the longitudinal sub-study	March, 2015	To investigate whether some factors in course design may affect FAV acquisition.	(1)What are the possible factors in teaching that may affect FAV development? (2)What are the possible factors in learning that may affect FAV development?
		To analyse the College English Syllabus for English Majors	March, 2015		
		To collect and analyse the course structure in the second semester of the longitudinal sub-study	September, 2015		
		To collect and analyse the course structure in the third semester of the longitudinal sub-study	March, 2016		

As it has been discussed in Section 1.2.2, these instruments can provide a clearer and more synthesizing evidence to answer the research questions.

3.4 Instruments and Data Collection

3.4.1 Instruments and Data Collection in the Longitudinal Sub-study

As it was introduced in Section 3.3, this research involves four sub-studies. In this section, the instruments used in each sub-study and the data collection procedure are

introduced. The first sub-study was the 12-month longitudinal study. In the longitudinal sub-study, the test instruments adopted were: (1) Receptive Vocabulary Levels Test (PVLТ); (2) Controlled Active (or Controlled Productive) Vocabulary Levels Test (CAVLT or CPVLT); (3) Lex 30 test; and (4) a composition of no less than 200 words. The reliability of PVLТ was tested by Read (1988) and Bayazidi and Saeb (2017), and both of them concluded that it was a reliable test. The reliability of CAVLT was tested by Laufer and Nation (1999), and they concluded that it was a reliable test. The reliability of Lex30 was repeatedly tested and confirmed by Walters (2012), Fitzpatrick and Meara (2004), and Fitzpatrick and Clenton (2010). Laufer and Nation (1995) tested the reliability of Lexical Frequency Profile (LFP) for compositions, and they concluded that the reliability was satisfactory.

It should be noted that FAV include words that are used spontaneously in both writing and speaking. In this part of research, however, only writing data were collected and analysed. It was believed that spoken data were not as efficient to measure L2 learners' FAV as written data were. According to Cobb (2003, p. 403), "Spoken language, especially conversation, does not require nuanced vocabulary since nuancing of meaning can be provided by shared context, deixis, facial expression, and so on. Most forms of writing, on the other hand, have greater need of nuanced vocabulary since written texts must be able to bridge gaps over space and time between unshared contexts." Therefore, in this part of research only written data were applied for FAV measurement. Two tests were adopted for the task. One was the Lex30 test, in which the testees were required to instantly write four response words to the given stimulus words. The other was a composition of no less than 200 words. The difference between the two tests is that the Lex30 test measures FAV out of contexts, whereas compositions measure FAV in contexts. Adopting both of the two

tests enabled the researchers to measure FAV from various perspectives, so that a more comprehensive portrait could be obtained.

The longitudinal study is considered to be a more appropriate choice than the cross-sectional study in this research. Although the cross-sectional study enables researchers to collect data in a relatively short period of time, it can be misleading if used to study developmental trends (Johnson & Christensen, 2014). In other words, if the study is to establish a time-over, the conclusion from cross-sectional studies tends to be weaker than that of longitudinal studies. By contrast, a longitudinal study occurs over time, which provides stronger evidence on developmental trends (Johnson & Christense, 2014). In this sub-study, a battery of vocabulary knowledge tests, including the passive vocabulary test (PVLТ), the controlled active vocabulary test (CAVLT), and the free active vocabulary tests (Lex30 and compositions) were conducted both in the first month and in the twelfth month.

The longitudinal sub-study in this research was a panel study, as the same individuals were studied for a period of time. The goal of a panel study is to understand how and why the panel members change with time passing by (Johnson & Christensen, 2014). By contrast, in a trend study, individuals are taken from a population to be studied over time, but each time the individuals may be different (Johnson & Christensen, 2014). As the sample kept the same over time, the panel study is believed to be stronger than the trend study (Johnson & Christensen, 2014). Meanwhile, this longitudinal sub-study was a prospective study, as it started in a present time and continued forward in time (Johnson & Christensen, 2014).

This section has provided an overview of the instruments and data collection in the longitudinal sub-study. More details on each of the instruments and on the data collection procedure are provided in Chapter 4.

3.4.2 Instruments and Data Collection in the Word Association Sub-study

The second sub-study, which was quantitative as well, was a word association study. In the word association sub-study, two types of instruments were adopted. One was Vocabulary Knowledge Scale (VKS) test, and the other was a word association test. As was discussed in Chapter 2, the word association approach is a well-established approach to investigate testees' lexical organisation in mental lexicon. The strength of word association test is that it is easy to be conducted, and it is time-efficient. However, testees' word associations to the target words tend to be affected by their level of familiarity to the words. Therefore, VKS tests were conducted to control the extraneous variable of familiarity.

In the sub-study, high-frequency words that tended to be frequently produced and that tended to be seldom produced by upper-intermediate to advanced L2 learners were selected. In the first month, the testees were tested with the VKS test and the word association test on words that were frequently produced. In the second month, the testees were tested with the VKS test and the word association test on words that were seldom produced by upper-intermediate to advanced L2 learners. The interval between the two tests was one month, so that reminding effects would be avoided. More details of the sub-study are provided in Chapter 5.

3.4.3 Instruments and Data Collection in the Interview

The third sub-study, which was qualitative, involved interviews. The sub-study was administered to achieve the fourth and fifth research objectives. The fourth research objective was to find possible factors in learning and teaching that may affect FAV development. The fifth research objective is to find effective ways to tackle the FAV

threshold phenomenon, if the threshold phenomenon is detected in the longitudinal sub-study.

All the questions in the interview were open-ended questions, as to answer the questions, the participants need to provide their own answers (Johnson & Christensen, 2014). By contrast, close-ended questions have predetermined answers, and participants need to select their answer(s). Open-ended questions were used in this sub-study, as the researchers need to know the interviewees' views in details, and the open-ended questions can provide researchers with more detailed information. The advantage of open-ended questions is that they can give researchers rich and in-depth information on the topic (Johnson & Christensen, 2014), and can help avoid misleading information and bias (Seidman, 1998). The interview used in this part of research was a standardized open-ended interview, as the questions had been written out in advance, and they would be read "exactly as written and in the same order to all the interviewees" (Johnson & Christensen, 2014, p. 234).

The interview was in a face-to-face form. Face-to-face interviews not only give researchers chances to have verbal communications with interviewees, but also provide them chances to observe interviewees. In other words, non-verbal communication and visual support can be provided in face-to-face interviews, which may bring better results (Neuman, 2004a).

This interview was also semi-structured. Different from non-structured interviews, questions in semi-structured interviews are prepared in advance, and all interviewees are treated in a like manner (Fontana & Frey, 2000). Different from structured interview, the semi-structured interview is more flexible, and it allows the researcher to ask questions in addition to prepared ones when necessary (Kvale & Brinkmann, 2015).

Fourteen questions were designed, which was believed to be at a manageable level to interviewees. Follow-up questions may emerge when more details were needed. The questions were designed surrounding four aspects, leading to four sets of questions. To ensure the validity of the interview, content-related evidence needs to be found. Content-related evidence refers to the evidence that can help researchers “judge the degree to which the evidence suggests that the items, tasks, or questions on your test adequately represent the domain of interest” (Johnson & Christensen, 2014, p. 142). According to Johnson and Christensen (2014), three questions should be asked to examine the content-related evidence: firstly, “do the items appear to represent the thing you are trying to measure?” Secondly, “have you excluded any important content areas or topics?” Thirdly, “have you included any irrelevant items?” (p. 142) Guided by the three questions, discussions were conducted between the researcher and the supervisor. In this way, the interview questions were examined and revised repeatedly. The final version of interview questions were believed to represent the four aspects, and it was believed that no important contents were excluded, and no irrelevant items were included in the interview. The interview questions are provided in Chapter 6.

The interview was conducted in June, 2016. The interviewees were interviewed by the researcher in a random order. The entire interview procedure was audio recorded with the interviewees’ consent. The researcher was aware that the interview settings might affect the interview result. According to Neuman (2004a, 2004b), the interview setting should make interviewees feel safe, relaxed, and confidential. Therefore, in order to get the best from the interviewees, the interview was carried out in a quiet place, and no third-party was present in the entire process. In the interview, adequate time was provided to the interviewees, so that they had enough time to think and express their

views without interruption. In addition, the interviewees were all volunteers to participate, as only those who agreed to participate would be interviewed. In other words, they were willing to be interviewed on the issue. Therefore, it was believed that interviewees answered the questions with a serious attitude, which was later evidenced by the interview recordings as well. If the interview was conducted again, most likely these interviewees would provide similar answers. Therefore, the validity of the interview was ensured.

3.4.4 Instruments and Data Collection in the Documentary Analysis

The fourth sub-study, which was qualitative as well, was documentary analysis. In the sub-study the documents, including the course structure that participants took in the period of the longitudinal sub-study and the Syllabus of English Courses for College English Majors, was collected and analysed. The documentary analysis was adopted in this research, as it may be the most achievable way to obtain the unit information needed in this research. Documents are a type of data that are frequently used in qualitative study (Johnson & Christensen, 2014). According to Johnson and Christensen (2014), documentary data usually fall into two categories: personal documents and official documents. The official documents are written, photographed, or recorded documents by some type of public or private organisations (Johnson & Christensen, 2014). As the course structure was made by school, it belonged to the official documents. The data in this sub-study belong to secondary data. Secondary data are data that have already existed (Johnson & Christensen, 2014). The reliability of the data is high, as the official documents are usually formal and accurate.

The research instruments and data collection in the longitudinal, word association, and interview sub-studies have all been tested in the pilot study. Details of the pilot study will be introduced in Section 3.5 and in the following chapters.

3.5 Pilot Study

Pilot studies were conducted prior to all the formal sub-studies in this research, except for the documentary analysis. Pilot studies can help researchers check whether the instruments or data collection methods are appropriate (Teijlingen & Hundley, 2002). The purpose to conduct pilot studies is to help researchers find possible weaknesses in research designs, so that failures of the study can be avoided at the very beginning. Therefore, pilot studies can help increase the likelihood to succeed in formal research (Teijlingen & Hundley, 2002).

In this research, a class of approximately 30 students was randomly selected from six classes for the pilot study. In this class, 11 students took the vocabulary knowledge test battery, another 11 took the VKS tests and word association tests, and the rest five students had interviews. Before the pilot study, information sheets were handed out to the participants, and the researcher invited them to participate in the study. The participants in the pilot study were also the year-three English-major students from the same university as the formal participants were; therefore, they shared the same characteristics as those participants who took part in the formal research. Results in the pilot study were only used to improve the formal tests and interviews. They were not used as formal data. After the pilot study, all the test papers were collected, and the participants were asked to keep the test content confidential. They were also reminded not to volunteer in the formal study that would be conducted later.

After the pilot tests and interviews were completed, the researcher had a discussion

with the participants. The research purpose and research design were explained to them again. Then the participants were asked to give comments on whether the tests and interview questions were clear enough to them, whether some additional questions were needed, and whether there were some irrelevant items that needed to be deleted. Participants' feedback, together with the discussions between the researcher and supervisors, played significant roles in the finalisation of research instrument design. It helped the researcher to decide the number of interviewees needed for the formal interview as well.

3.6 Sampling and Participants

3.6.1 Sampling and Participants in the Longitudinal Sub-study

The participants were sampled from year-three university students, and their major was English. All these students had experienced standard elementary and middle school education in China, indicating that they had already been studying English for at least six years before they entered the university. Therefore, students in the sampling pool had been studying English for approximately nine years at the beginning of this part of research, and around 10 years at the end of it.

The university that was chosen to conduct the study was a public university in China, thus students studying at this university had passed the competitive nation-wide university entrance examination, in which English was one of the five subjects. At the beginning of this part of research, they were almost at the end of their year-three university study. Having English language as their major, these students had an intensive course in English, which offered English training in various perspectives, such as reading, writing, speaking, western culture, grammar, etc. The course

structure was designed on the basis of the College English Syllabus for English Majors by the Chinese Ministry of Education in 2000; thereby students would take ten to fifteen hours' English classes each week. In the four years, the units students took were listed below:

- Intensive Reading
- Introduction to Linguistics
- British Literature and Anthology
- Theory and Practice of Translation
- Phonemic
- Selected Reading in English Essays
- American Literature and Anthology
- Selected British and American Newspaper Reading
- English Academic Thesis Writing
- Research methods in English Teaching
- Phonetics
- English Speaking
- English Listening
- English Speech
- Introduction to Western Culture
- Introduction to English Literature
- Lexicology
- Pragmatics
- Advanced English Writing
- Audio-visual English
- English Writing

- English Pedagogy
- Overview of English-speaking Countries
- English Grammar

All of these units were taught in English, and most teachers who taught the units were Chinese teachers with master's degrees in foreign language study, such as linguistics, applied linguistics, literature, and translation. A few teachers teaching in the department had obtained doctor's degree in these areas. English Speaking, and sometimes Introduction to Western Culture and Advanced English Writing as well, were mainly taught by teachers whose native language was English.

There was no unit in the course that was specifically designed to assist word acquisition. Although Lexicology was provided in year three, this unit primarily dealt with vocabulary in a linguistic point of view. The content of the unit was primarily on English words' origin, their structure, history, development, and formation. New words were usually learned explicitly in Intensive Reading, or acquired implicitly in other units. In addition to studying the language from formal university instruction, most of the students would try to have more English learning out of class, such as from English TV and radio programs, and from the Internet; they also sought chances to have oral practice with native-speakers, through English Corner, for instance, which was usually a weekend activity for oral English practice with others.

Most of the students would seek discipline-related work after they graduate from the university. For example, they may become English teachers, translators or staff in companies doing international businesses. It thus seemed reasonable to assume that most of them have a relatively high level of motivation to study the language. Taking all the above factors into consideration, it was concluded that the sampling pool represent the upper-intermediate to advanced L2 learners in China.

The sampling method adopted in the longitudinal sub-study was convenience sampling. In convenience sampling, people who are available, who can be easily accessed, or who are willing to volunteer are included in the sample (Johnson & Christensen, 2014). Convenience sampling is different from random sampling, as in convenience sampling, not everyone belonging to the population will have equal chances to be included in the sample. Technically speaking, a population cannot be generalized from a convenience sample (Johnson & Christensen, 2014). However, in reality the majority of experimental researchers tend to use convenience samples instead of random samples, due to practical constraints or ethical considerations (Johnson & Christensen, 2014). In this part of research, convenience sampling was adopted, as according to the ethics requirement, only those who were willing to participate in the research could be included in the sub-study. Before the longitudinal sub-study was conducted, the researcher called on students to participate in it. Only those who expressed their willingness to do the tests and signed the consent form would be recruited.

It should be noted that when convenience sampling is adopted, researchers should describe the characteristics of the participants, and decide whom the participants may represent (Johnson & Christensen, 2014). It was believed that the convenience sampling in this part of research did not undermine the representation of the target population, which was the upper-intermediate to advanced L2 learners in China. There were 27 volunteers who took part in the sub-study, and all the volunteers shared the characteristics with the rest in the sampling pool. The reason why they were more willing to participate in this part of research than their colleagues may most likely be because they felt more interested in it, or they wanted to test their vocabulary sizes by using this sub-study as an opportunity.

The average age of the participants in the 12-month longitudinal sub-study was approximately 20 years old, and at the end of the longitudinal sub-study, the average age of the participants was around 21. Among the 27 participants, there were 21 (77.8%) females and 6 (22.2%) males. The gender imbalance of the subjects was normal, owing to the fact that female students to a great extent outnumbered male students in the department.

3.6.2 Sampling and Participants in the Word Association Sub-study

The participants of this sub-study were sampled from the same sampling pool as the participants of the longitudinal sub-study. The word association sub-study adopted convenience sampling as well, in which only students who were willing to do this part of research participated in it. The convenience sampling did not see significant impact on participants' representation of the upper-intermediate to advanced L2 learners, as participants shared the same key features with the target population in education background and motivation to study English.

In total, 30 participants took part in the word association sub-study. The participants were year-three undergraduate students, whose major was English as a foreign language. The average age of the participants was around 20. Among the participants there were 23 females, which accounted for 76.7% in the sample; and 7 males, which accounted for 23.3% in the sample. Again, the gender imbalance of the subjects was normal, owing to the fact that female students to a great extent outnumbered male students in the department.

3.6.3 Sampling and Participants of the Interview

Three participants were involved in the interview, including one male and two females. The reason why no more interviewees were interviewed is that the interview is considered to be a supplement to the word association sub-study. In the word association sub-study, the cognitive reasons for FAV stagnation were investigated. As is discussed in Chapter 1, this research assumes that underlying the effective teaching and learning methods, it is the cognitive mechanism that works in FAV development. In other words, the cognitive mechanism behind those methods is the root cause for FAV growth. For this reason, it is suggested that more attention should be paid to the mental lexicon construction in FAV development. Guided by the findings of the mental lexicon construction, effective teaching and learning methods may be identified more effectively, and new teaching and learning strategies may be designed. The interview in this research serves as a supplement for the word association sub-study. The effective learning and teaching methods that the interviewees felt effective were cross-compared with the findings from the word association study, and to check whether those strategies can be explained by findings from the word association study. Results of the pilot interview also showed that three interviewees were enough, as the strategies interviewees reflected were limited and similar, and their reflection on the attention to FAV teaching and learning was relatively consistent. After three interviewees no new themes came out.

The participants were sampled from the same sampling pool as participants of the longitudinal sub-study and the word association sub-study were. Like the two previous sub-studies, convenience sampling was adopted due to ethical considerations. Before the interview was conducted, the researcher called on students to participate in

it. Only those who showed interest and willingness to do the interview would be recruited. The consent form for the participation was signed by participants as well. The time to conduct the interview was in June, 2016. At the time of the interview, the interviewees were around 21 to 22 years old, and they were seniors in the second semester. They had completed all the courses required for the 4-year university study. In addition, they had just completed a thesis writing of about 5,000 words, which was a compulsory requirement for graduation. The topics of the thesis were required to be related to L2 linguistics and applied linguistics, literature, translation, or culture. All the interviewees participated in TEM8 exam in March, 2016. TEM8 is a national test of English as a foreign language in China. The full name of TEM is Test for English Majors. TEM has two levels, TEM4 and TEM8. Certificates of TEM4 and TEM8 are important to English-major students, as they are usually required by recruiters in English related jobs. Meanwhile, TEM tests, especially TEM8, have relatively high ESL requirements to testees, and are relatively difficult to pass. The high level of difficulty can be evidenced by the average national passing rate of TEM4 being 49.92% (Renmin University of China, 2017), and that of TEM8 being 40.60% (Tongji University of China, 2017) in 2016. In order to pass TEM8, students usually spend a great amount of time to prepare for it. In addition, all of the three interviewees participated in the postgraduate entrance examination in January, 2016, in which English was one of the subjects.

3.7 Data Analysis

There were two types of data in this research: numerical data from quantitative sub-studies, and textual data from qualitative sub-studies. The analysis methods for both of the two types of data are introduced in this section.

3.7.1 Quantitative Data Analysis

The Statistical Package for Social Science (SPSS) software was adopted to analyse the data obtained from both the longitudinal sub-study and the word association sub-study. The data that were processed by SPSS included the scores of passive vocabulary test, controlled active vocabulary test, Les30 test, LFP from compositions, and word association tests. The Vocabulary Knowledge Scale (VKS) test, which was a Likert-scale test used to measure testees' familiarity level with the cue words in the word association sub-study, was analysed by SPSS as well. The statistical analysis methods included Descriptive Statistics, Paired-sample T test, and One-way Repeated Measures ANOVA. Details of the statistical analysis for each sub-study are provided in the following chapters. SPSS was considered to be an appropriate tool, as it met all the calculation needs in the two studies. In the word association sub-study, the software AntConc 3.5 was adopted as well. AntConc 3.5 was used to count the frequencies of candidate cue words in the word association tests. Details of the quantitative data analysis are provided in Chapter 4 and Chapter 5.

3.7.2 Qualitative Data Analysis

The first qualitative sub-study in this research was the interview. The analysis of the interview was guided by the constructive grounded theory. According to Charmaz (2006), the constructivist grounded theory “consists of systematic, yet flexible guidelines for collecting and analysing qualitative data to construct theories ‘grounded’ in the data themselves” (p.2). Based on the constructivist grounded theory, the data analysis in this sub-study took a series of steps. Firstly, the entire process of the interview was recorded. Then the interview was listened to by the researcher, and was transcribed by the researcher. Next, the transcripts were rechecked by the

researcher to ensure that every word had been transcribed correctly. In the next step the transcript was segmented. Units in the transcript that contained valuable information were found and marked. Then coding was conducted. The researcher read the transcript line by line, and put descriptive phrases as codes beside each segment. In the sub-study, both pre-existing and inductive codes were used. The pre-existing codes were designed in advance based on the research objectives of the interview. For example, the pre-existing code of “effective learning strategies” was made to capture learning activities that may be effective in FAV learning; the pre-existing code of “effective teaching strategies” was made to detect teaching methods that may be helpful to FAV learning. Inductive codes were not prepared in advance, but came into being when the researcher found some unexpected valuable information from the data. For example, one of the inductive codes in this sub-study was “the number of learning strategies used by the interviewee”, which was valuable information to judge whether students adopted a variety of activities in FAV learning. After the coding was completed, the codes were re-examined by the researcher and the supervisor. When there was inconsistent coding, discussions were made until agreement was achieved. In this way the intercoder reliability and intracoder reliability was ensured. All the codes were put into a master list. Then the codes were enumerated and categorised. The researcher systematically analysed the categories and codes, and attempted to construct themes and theories that were grounded in them.

3.8 Conclusion

This chapter serves as a methodological foundation for the entire research project. It first introduces the research aim and objectives, then introduces the research method used in this study. Information on the research instruments, participants, sampling

process, pilot study, and data analysis methods is provided as well. In the next chapter, the first quantitative sub-study, the 12-month longitudinal sub-study, is presented and discussed.

Chapter 4

The Developmental Process of Free Active Vocabulary

4.1 Introduction

This chapter explores the developmental process of free active vocabulary (FAV) in a 12-month period of time. Moreover, it investigates the relationship between FAV and other types of vocabulary, i.e., the passive vocabulary (PV) and the controlled active vocabulary (CAV). The chapter first poses research questions, and then presents research method, which includes instruments, data collection, data preparation and data analysis. At the end of the chapter, results are reported, and answers to the research questions are discussed.

4.2 Research Questions

The questions that are explored in this chapter are:

- What is the development pattern of FAV when L2 learners reach upper-

intermediate to advanced proficiency level?

- If the threshold phenomenon occurs in the FAV development, what are the features of the threshold phenomenon?
- What is the developmental process of PV when L2 learners reach upper-intermediate to advanced L2 proficiency level?
- What is the relationship between FAV and PV?
- What is the developmental process of CAV when L2 learners reach upper-intermediate to advanced L2 proficiency level?
- What is the relationship between FAV and CAV?

4.3 Research Method

This section presents the major components of the research method adopted in the sub-study, including instruments, procedure of data collection, data preparation, and data analysis.

4.3.1 Instruments

The participants in this longitudinal sub-study have been introduced in Chapter 3. The instruments adopted in the sub-study included three vocabulary tests and two writing tasks, which are detailed and explained in this section.

4.3.1.1 Test of Receptive Vocabulary

The test of receptive vocabulary adopted in this sub-study was the Vocabulary Levels Test (VLT). It should be noted that the Vocabulary Levels Test (VLT) in fact has two types. The first type is for measuring passive vocabulary, and the second type is for measuring controlled active vocabulary. Therefore, in order to be more accurate, it is proposed in this research that the two types of Vocabulary Levels Test be renamed as

Passive Vocabulary Levels Test (PVLТ) and Controlled Active Vocabulary Levels Test (CAVLT), according to their different functions. In PVLТ, there are five sections: Section A for words from the first 2,000 (i.e., 1-2,000) word frequency band, Section B for words from the 3,000 (i.e. 2,001-3,000) word frequency band, Section C for words from the 5,000 (i.e. 4,001-5,000) word frequency band, Section D for words from the 10,000 (i.e. 9,001-10,000) word frequency band, and Section E for academic words.

The original Passive Vocabulary Levels Test was designed by Nation (1983, 1990), and its reliability was tested and verified by Read (1988). The PVLТ adopted in this research was the revised and expanded version designed by Schmitt et al. (2001). In the early version of PVLТ by Nation (1983, 1990), one-third of the tested words in Section A are sampled from the first 1,000 (i.e. 1-1,000) word frequency level, and the rest two-thirds of the tested words in Section A are sampled from the second 1,000 (i.e. 1,001-2,000) word frequency level. In the revised version designed by Schmitt et al. (2001), half of the tested words in Section A are sampled from the first 1,000 word frequency level, and the other half are from the second 1,000 word frequency level. Other sections are similar to those of the original PVLТ. The tested words in Section B are from the 3,000 word level, i.e., they are sampled from the 2,001-3,000 word frequency band. Similarly, in Section C the tested words are from the 5,000 word level, i.e., the 4,001-5,000 word frequency band, and in Section D the tested words are from the 10,000 word level. i.e., the 9,001-10,000 word frequency band. The word frequency bands are constructed primarily on the basis of the General Service List (GSL). Schmitt et al. (2001) devised three revised versions of PVLТ, and tested their validity. In these revised versions, the tested words in the academic section are from the Academic Word List (AWL). In the original version by Nation (1983, 1990),

however, the tested words in this section are from the University Word List (UWL), as there was no AWL yet at the time when it was compiled. All these versions of PVLT are in the form of matching tests. In the revised PVLT, each section has 10 clusters, as compared to 6 clusters in the original version by Nation (1983, 1990). Each cluster is composed of 3 tested words on the right column, and 6 paraphrases on the left column. An example from Schmitt et al. (2001) is shown below:

1 copy	
2 event	_____ end or highest point
3 motor	_____ this moves a car
4 pity	_____ thing made to be like
5 profit	another
6 tip	

Since the revised PVLT contains 5 sections, and each section contains 30 test items, the test has 150 test items in total. In each section, the ratio of noun, verb, and adjective is 3:2:1. According to Schmitt et al. (2001), the average time needed to take the test is 31 minutes.

PVLT has numerous strengths. It is easy and quick to be conducted and scored, and need no special equipment. It can provide researchers and teachers with a relatively comprehensive picture on test takers' passive vocabulary knowledge, since the test contains various sections to test words from different frequency levels. Compared with the original version of PVLT by Nation (1983, 1990), the revised versions by Schmitt et al. (2001) are more updated, as they adopted AWL instead of UWL to sample tested words in the academic word section. The revised versions also sample tested words in a more balanced and rational way. As has been introduced, in the early version of PVLT by Nation (1983, 1990), one-third of the tested words in Section A are sampled from the first 1,000 (i.e. 1-1,000) word frequency level, and the rest two-thirds of the tested words in Section A are selected from the second 1,000 (i.e. 1,001-

2,000) word frequency level. In the revised version designed by Schmitt et al. (2001), half of the tested words in Section A are sampled from the first 1,000 word frequency level, and the other half are from the second 1,000 word frequency level. Moreover, the revised versions cover more tested items than the original version. There are 18 test items in each section in the original version. By contrast, there are 30 test items in each section in the revised version. Bayazidi and Saeb (2017) tested the reliability of two of the three revised versions by Schmitt et al. (2001), and they concluded that the two are reliable tests. Due to the strengths of the revised PVLTL, one of the revised versions was adopted in this sub-study. The reliability of the adopted version was tested and proved by Bayazidi and Saeb (2017).

It should be noted that only one version of PVLTL by Schmitt et al. (2001) was adopted in this longitudinal sub-study. In the first month and the twelfth month, the test was conducted twice, and then the test results were compared. It was due to the consideration that although Schmitt et al. (2001) tested and proved the revised versions' validity, they did not test whether those versions were parallel to each other. In addition, as the interval between the two tests was as long as 12 month, the possibility that the first-month test would have a great reminding effect on the twelfth-month test was low. Therefore, only one version was adopted. In 2017, Bayazidi and Saeb (2017) tested two of the three revised PVLTL versions by Schmitt et al. (2001), including the one that was applied in this sub-study. He found although both of the two versions were highly reliable tests, they were not equivalent or parallel. Therefore, the two should not be treated as equal forms and used as parallel tests, particularly in longitudinal studies (Bayazidi & Saeb, 2017). Appendix A presents the PVLTL version adopted in this sub-study.

4.3.1.2 Test of Controlled Active Vocabulary

The structure of CAVLT is parallel to that of PVLTL. The CAVLT contains 5 sections as well, i.e., Section A for the 2,000 word frequency level, Section B for the 3,000 word frequency level, Section C for the 5,000 word frequency level, Section D for the 10,000 word frequency level, and Section E for the university words. In total, CAVLT contains 90 test items, with each section having 18 test items. CAVLT is in the form of cued recall. In the test, full sentences with blanks for the target words are provided. The first two letters of the target words are provided as well.

It is necessary to note that in the revised PVLTL by Schmitt et al. (2001), which was adopted in this sub-study, the tested items in the academic section (Section E) are sampled from the Academic Word List (AWL). However, in CAVLT tested items in the academic section (Section D) are selected from the University Word List (UWL). Therefore, it is necessary to investigate the similarities and differences between UWL and AWL. The UWL, which was designed by Xue and Nation (1984), is a list of words that are used in academic texts with various subjects. The list contains 836 words. These 836 words are the most frequently used academic words that do not belong to the first 2,000 words in the General Service List (GSL) by West (1953). The AWL is a new version of UWL, and it was designed by Coxhead (2000). Coxhead (2000) holds that the first 2,000 words in GSL are basic words that any ESL learners need to grasp before studying the academic vocabulary. Therefore, just like UWL, AWL does not contain the first 2,000 words from GSL. AWL contains 3,100 words belonging to 570 word families which originate from the 836 words of UWL. According to Schmitt et al. (2001), AWL gives better coverage of academic texts while listing fewer words than UWL, as AWL presents academic words in form of word families. Different from other sections in PVLTL and CAVLT, the academic

section in PVLТ and CAVLT is not primarily frequency driven. In fact, the AWL and UWL contain words of various frequency levels above the 2,000 frequency level. In other words, the academic section in the original and revised PVLТ and CAVLT contains words of frequency levels from 2,000 to 10,000 (i.e. 2,001-10,000). The reason why UWL is adopted in CAVLT but AWL is adopted in the revised PVLТ is because AWL was not compiled yet when CAVLT was devised. However, since AWL was devised based on UWL, the two are basically consistent to each other.

Laufer and Nation (1999) tested the validity and reliability of CAVLT. They concluded that it is a reliable and valid test, and just like PVLТ, the advantage of CAVLT is that it is easy to conduct and score. The version of CAVLT adopted in this sub-study is presented in Appendix B.

4.3.1.3 Test of Free Active Vocabulary

4.3.1.3.1 Lex30 Test

One of the tests conducted to measure the testees' FAV in this sub-study is the Lex30 test, which was designed by Meara and Fitzpatrick (2000). Lex30 is in the form of word association tasks. However, by nature it is not a word association test, as its score is based on the frequencies the associations belong to, rather than association types they are of. There are 30 cue words in Lex30, and testees are required to produce four responses to each of the cue words if they can. All the 30 cue words are sampled from the first 1,000 frequency band. It is believed that words from that frequency level can be recognised by test takers of various language proficiencies, and is therefore suitable to be used as cues for a wide range of testees (Meara & Fitzpatrick, 2000). In addition, the 30 cues are believed to be able to elicit responses that do not fall into the first 1,000 frequency level, and do not tend to elicit strong

primary associations (Fitzpatrick & Clenton, 2010) as well. Lex30 is easy and time-saving to be conducted, as it only takes 15 minutes to complete (Fitzpatrick & Clenton, 2010). Moreover, it gives testees chances to produce responses more freely, without the confine of context (Meara & Fitzpatrick, 2000). The validity and reliability of Lex30 has been repeatedly tested and proven (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004; Walters, 2012). The Lex30 test adopted in this sub-study is presented in Appendix C.

4.3.1.3.2 Writing Tasks

Although writing tasks are not as time-saving and convenient to be carried out as Lex30 is, both of them were conducted to measure FAV in this sub-study. The first reason is that although Lex30 possesses many strengths in measuring FAV, it is believed by some scholars that “producing response words to stimuli...is not the same as spontaneously producing words in a composition, of course” (Siok & James, 2006, p. 297). In other words, measuring FAV in context may offer researchers more accurate information on FAV, although it exerts on testees a limit of context in word production. The second reason is that although Meara and Fitzpatrick (2000)’s original Lex30 test can show scores of FAV at various frequency levels by using a software, the software seems not available to other researchers. The Internet version of Lex30 can generate a general score, but cannot generate scores for responses at different word frequency levels. According to the scoring standard of Lex30 (Meara & Fitzpatrick, 2000), the test result of Lex30 can only reflect the number of responses that fall out of proper nouns, numbers, high-frequency function words, and the first 1,000 most frequent words. More detailed information on FAV cannot be obtained from it. The writing tasks, however, may offer researchers more detailed information

on FAV. Owing to the consideration that both Lex30 and writing tasks have strengths and weaknesses in evaluating FAV, to get more accurate and comprehensive information on it, both of them were adopted in this research as measuring tools.

As has been pointed out in Chapter 2, there is no instrument that can directly measure the size of FAV. The size of FAV therefore is usually expressed by lexical richness. It is believed that if the lexical richness is high, then the test taker may have more FAV available in his lexical repertoire (Laufer, 1991). There are two ways to measure lexical richness. One is by Lexical Frequency Profile (LFP), and the other is by four indexes including lexical variation (LV), lexical originality (LO), lexical sophistication (LS), and lexical density (LD). The first method, the Lexical Frequency Profile (LFP), measures the richness of a text by calculating how many words in the text belong to different frequency categories. LFP usually includes four categories. The first category, which is called Band 1, contains the first 1,000 frequent words in English; the second category, Band 2, contains the second 1,000 frequent words in English; the third category, Band 3, contains the academic words; and the fourth category, “Not in the lists”, are words that do not belong to any of the 3 categories mentioned above. LFP shows the distribution of words (including tokens, types and word families) of a text in the four categories. The computer program that conducts LFP calculation is Vocabprofile, which was later updated to be RANGE. In Vocabprofile, UWL was adopted to be Band 3, the academic words. In RANGE, AWL was adopted to be the academic words.

RANGE was developed by Paul Nation, Alex Heatley, and Averil Coxhead at Victoria University of Wellington. It can be downloaded from Nation’s website as part of the website of the Victoria University of Wellington (Heatley, Nation, & Coxhead, 2002). RANGE contains three word lists. The first word list (BASEWRD1) contains

the first 1,000 frequent word families of English; the second word list (BASEWRD2) contains the second 1,000 frequent word families; and the third word list (BASEWRD3) is the Academic Word List. Both BASEWRD1 and BASEWRD2 are compiled based on the General Service List. All the three lists contain words' base forms, derived forms, and inflectional forms. Results by RANGE show the numbers and percentages of tokens and types that fall into each of the three word lists, as well as the numbers and percentages of tokens and types that are not in the three lists. The number of word families that belong to the three lists and those that are not in the lists are counted as well.

The second way that is widely used to measure lexical richness is by lexical variation (LV), lexical originality (LO), lexical density (LD), and lexical sophistication (LS). Laufer (1991) may be one of the earliest scholars using these four indexes to measure FAV size. According to Laufer (1991), LV is the type/token ratio, the formula of which is $LV = (\text{number of types} \div \text{number of tokens}) \times 100\%$. LV indicates how testees incline to repeat the same words in writing. The higher the LV, the more varied the words are in writing. According to Laufer (1991), LD refers to the percentage of lexical words in the text. The lexical words are also called content words, which include nouns, verbs, adverbs, numerals, and adjectives. Information is usually conveyed by lexical words. Functional words, however, do not convey as much information, and they are usually used to show relations between words and sentences (Thoughtco, 2017). Functional words include conjunctions, prepositions, articles, and auxiliaries (Thoughtco, 2017). Since a word is either a lexical word or a functional word, the total number of tokens can be calculated by adding the two types of words together. The formula for calculating LD is $LD = (\text{number of lexical words} \div \text{number of tokens}) \times 100\%$. The higher the LD is, the "denser" the text is, since there are more

lexical words in the text to convey information. According to Laufer (1991), LO is the percentage of unique words that are used by a writer but not by other writers in the group. The higher the LO is, the more words there are that can be produced by the testee than by his peers. The formula of LO is $LO = (\text{number of lexemes unique to one writer} \div \text{total number of tokens}) \times 100\%$. According to Laufer (1991), LS refers to the percentage of “advanced words” used in writing. There is no set standard to judge which words are “advanced words”. Instead, the judgment of “advanced words” is flexible, depending on the testees’ proficiency level. For example, in Laufer (1991)’s study, the participants were year-one university students in English language department; therefore, the researcher decided that “advanced words” were those from UWL. He pointed out that for testees who were school learners, “advanced words” could be those taught in upper grades.

All the four indexes of LD, LV, LO, and LS may be appropriate tools to measure the quality of a piece of writing, but they seem to be problematic as an indicator of FAV size. The problem of LV and LD is that the two seem to ignore the factor of word frequency in calculation. According to Schmitt et al. (2001), words are usually learned in layers. Specifically, “learners acquire more frequently used words before they acquire less frequently used ones” (Schmitt et al., 2001, p. 67). Testees may need to learn more words in higher frequencies before they can grasp more words in lower frequencies. Therefore, if more FAV fall into lower frequencies, it may indicate that the testee has a bigger FAV repertoire. However, the LV and LD do not consider LFP in their calculation, which may make the two not sensitive to FAV size. For example, two testees are asked to write a composition of 200 tokens. The first testee produces 60 different types, and all the 60 types are from the first 2,000 frequent words; the second testee produces 60 different types as well, but 40 of them are from the words

beyond the first 2,000 frequent words. According to the formula $LV = (\text{number of types} \div \text{number of tokens}) \times 100\%$, the two testees have the same LV value. However, it may not be rational to draw a conclusion from this result that the two testees have the same FAV size. More likely, they have different FAV repertoires, since the second testee can produce more lower-frequency words than the first testee. LD shares the same deficiency. For instance, two testees are asked to write a composition of 200 tokens, and in both of their writings there are 80 lexical words. All the first testee's lexical words are from the first 2,000 frequent words, and 60 of the second testee's lexical words are from the words beyond the first 2,000 frequent words. According to the formula $LD = (\text{number of lexical words} \div \text{number of tokens}) \times 100\%$, the two compositions result in the same LD value. But the two testees may have different FAV size, since the majority of their lexical words are from different word frequency bands. Therefore, LV and LD seem to be inappropriate measures of FAV size. LO and LS to some degree share the same idea as LFP in that they all indicate that words at lower frequencies may convey important information on FAV size. In LO, lower-frequency words are represented by "unique words"; in LS, they are expressed by "advanced words". However, both "unique words" and "advanced words" are problematic in that there are no set standards for them. This weakness may make researchers' judgment subjective and arbitrary, and therefore make it hard to have cross-comparisons of research findings. In addition, the list of "unique words" and "advanced words" has to be compiled by researchers themselves, which is not time-efficient.

Due to the weaknesses of LO, LD, LV, and LS in FAV evaluation, it is proposed in this research that LFP be adopted to evaluate FAV size. Laufer and Nation (1995) tested the reliability and validity of LFP, and they concluded that the reliability and

validity is satisfactory.

In this sub-study, two topics were selected for students' writing tasks. Both of the two topics were referenced from IELTS, but minor changes were made according to the purpose of the sub-study. The length of writing was changed to be no less than 200 words, so that the work load of testees was rational, meanwhile enough tokens could be obtained for reliable LFP results (Laufer & Nation, 1995). The two topics were at the same difficulty level. Both of them were of a discussion type that could elicit testees to express their own opinions. In this way, primary responses could be avoided. Topics that seemed to be too easy, such as those on food, entertainment, and travelling, were not chosen, owing to the fact that those topics may elicit mostly high-frequency words. The two topics were presented in Chinese, so that testees' FAV production would not be affected by the given information. The topic in the first-month test was on factors that could help people achieve success, and the topic in the twelfth-month test was on factors that could help a country develop. Detailed information on the two writing tasks is presented in Appendix D and E.

4.3.2 Data Collection Procedures

4.3.2.1 Pilot Study

A pilot study was administered before the formal data collection. 11 year-three students who were studying in the same major and from the same university took part in it. They were asked to do the PVLTL, CAVLT, Lex30, and a writing task with one of the two topics in the formal tests. The pilot study aimed to ensure that the test instruction and test items were clear enough to test takers, and to find the suitable length of time for testees to complete each test. Based on the results and testees' feedback, test instructions were changed to be in Chinese. The pilot study showed that

some participants were able to spell the tested words in CAVLT, but they did not pay adequate attention to grammar. Therefore, in the instruction of CAVLT, a reminder was added. The reminder was: “Please pay attention to grammar such as number, tense, person, and possession. Please write the target words with correct grammatical forms.” Moreover, in order to assist test takers to choose the right word to spell in CAVLT, Chinese translation of the target words was added after the blanks. It was found that some participants did not notice that on the other side of the paper there were also test items. Therefore, at the end of each page a short note was provided, saying that there were more test items on the other side of the paper. The length of time for each test was decided as well. There would be 30 minutes for PVLTL, 30 minutes for CAVLT, 15 minutes for Lex30 test, and 40 minutes for the writing task. Before the writing task, there would be a 5-minute break. Therefore the formal tests would take 120 minutes in total.

4.3.2.2 Formal Data Collection

Due to the fact that the sub-study is a longitudinal one of 12 months, formal data collection was conducted twice, with the first one being conducted in May, 2015, and the second one being conducted in May, 2016. May was considered to be a suitable time to collect data, since it was in the middle of a semester, and students were in their normal track of study at the time. Before the data were collected, instructions for PVLTL, CAVLT, Lex30 and the writing tasks were orally presented to the testees by the student investigator. Extra time was offered to test takers to read the written instructions as well. If they had any doubts on the test format or requirements, they were encouraged to raise questions. The testees were reminded on the length of time allotted to each test, and they were suggested to manage their time according to it.

Then test papers were handed out to the testees. When there was 5 minutes left for each test, the testees were notified on it, and they were suggested again to manage their time. It was noticed that most of them could complete the tests on time. In both the first data collection and the second one, the procedure was the same. After the tests were completed, all the test paper was collected. Since the interval between the first and second data collection was 12 months, it was assumed that the interval time was long enough, therefore the reminding effect was minimal. In addition, no results were released to the participants after the first-month test, so the possible influence of the first-month test on the twelfth-month test was minimal. In the PVLТ test, the investigator repeatedly reminded testees that if they did not know which paraphrase a tested word should be matched with, they should not guess and should just leave the blank empty. The purpose was to ensure that results of PVLТ were not affected by guessing. The detailed information on the tests is demonstrated in Table 4.1.

Table 4.1. *Tests Adopted in the Longitudinal Sub-study*

Test type	Activity	Purpose	Time to conduct the test
The Passive Vocabulary Levels Test	Match the tested words with their equivalent paraphrases	To measure testees' passive vocabulary size	May, 2015 May, 2016
The Controlled Active Vocabulary Levels Test	Complete the tested words according to the given context	To measure testees' controlled active vocabulary size	May, 2015 May, 2016
Lex30 test	Instantly write down four responses that come into the testees' mind when encountering the cue word	To measure testees' free active vocabulary size	May, 2015 May, 2016
Writing Task One	Write a composition of no less than 200 words on the given topic	To measure testees' free active vocabulary size	May, 2015
Writing Task Two	Write a composition of no less than 200 words on the given topic	To measure testees' free active vocabulary size	May, 2016

4.3.3 Data Preparation

4.3.3.1 Selection of Valid Data

Test papers were collected after the testees completed the tests. Then valid data were selected. Since the sub-study was longitudinal, test papers done by the testees who failed to participate twice were removed. Only the data done by subjects who participated in both the first-month and the twelfth-month tests were kept for further selection. Data with handwritings that were difficult to be recognised, and data that were suspected to be done carelessly were removed. In total six participants' data were removed. After the selection, 21 subjects' test papers were determined as valid data for final analysis.

4.3.3.2 Scoring of PVLТ

In PVLТ test, if a tested word is matched with the right paraphrase, one point will be given. Since there are 30 tested words in each section, and there are 5 sections in the test, the total score of PVLТ is 150 points, with each section having 30 points.

4.3.3.3 Scoring of CAVLT

The scoring program for CAVLT can be found on website of Lextutor (2017). Following Laufer and Nation (1999), data were closely checked and necessary revisions were made before they were put into the computer. Minor mistakes in spelling, i.e., mistakes that did not affect word recognition (such as “concieve” for “conceive”), were corrected. Grammatical mistakes were corrected as well. All other data were kept unchanged and typed into the computer. The computer program then produced scores for CAVLT by hundred-marks. There were 6 scores for each participant, with 5 scores for the 5 sections in CAVLT, and a total score for the whole

test.

Since in CAVLT grammar mistakes and minor spelling mistakes are corrected before scoring, the CAVLT is more of a test measuring the breadth of controlled active vocabulary, rather than measuring the depth of it (Laufer & Nation, 1999). From the hundred-mark score, the controlled active vocabulary size can be estimated. The test items at each section (except the academic word section) are sampled from 2,000 word frequency level, 3,000 word frequency level, 5,000 word frequency level, and 10,000 word frequency level, and each of these levels contains 1,000 word families. Therefore, if nine out of the 18 items in Section A are answered correctly, then the score of Section A is 50 points. The score indicates that 50% of the 1,000 word families at that frequency level may be known by the testee. The controlled active vocabulary size at that frequency level is therefore around 500 word families. In other words, there are around 500 word families at the level that are readily available for the testees' controlled productive use. The section of academic words contains items sampled from 836 words frequently used in academic texts. The size of a testee's academic controlled active vocabulary therefore can be estimated according to the percentage of correct answers in the section as well.

4.3.3.4 Scoring of Lex30 Test

According to Meara and Fitzpatrick (2000), the scoring of Lex30 is conducted by a software similar to Vocabprofile. In Lex30, associated words that belong to the Level 0 word list, i.e., proper nouns, numbers, and high-frequency function words, get zero points. Associated words that belong to the Level 1 word list, i.e., the first 1,000 frequent content words in English, are assigned zero point as well. Any association that does not belong to these two categories would obtain one point. Since there are in

total 30 cues in the test, and testees are required to have 4 associations to each cue, the maximum score a testee can obtain is 120 points.

Before being entered into the computer for scoring, responses were processed. Firstly, like in CAVLT, words with minor spelling errors that did not affect recognition (e.g. “concieve” for “conceive”) were corrected. Words that had evident spelling mistakes were kept unchanged. Secondly, if one cue elicited a response but the response was written more than once by the testee, only one response was kept. All others were deleted, as the program would score the response repeatedly if it appears repeatedly. Thirdly, following Meara and Fitzpatrick (2000), words with inflectional suffixes, including plural (e.g. apples, eggs), third person singular present tense (e.g. talks, runs), past tense (e.g. talked, ran), past participles (e.g. written, spoken), present progressive (e.g. working, running), comparative (e.g. better, bigger), superlative (e.g. most, best) and possessive (e.g. his, its), were lemmatised, and they were considered the same as their base forms. For example, the response word “digs” to the cue word “dig” would be lemmatised to be its base form “dig”. Since the response “dig” was the same as its cue, it would score zero point. For another example, if the cue word “dig” elicited responses of “hole” and “holes”, the responses of “hole” and “holes” would together score one point. This is because “holes” should be lemmatised as “hole”, and “hole” belongs to the second 1,000 frequency band. Meara and Fitzpatrick (2000), who designed the Lex30 test, holds that some affixes, usually those most frequent and regular derivational affixes, should be lemmatised to be base forms as well. These derivational affixes include -able, -er, -ish, etc. However, in this research the principle was not followed, due to the consideration that ESL learners may not have as much language instinct and grammar knowledge as native speakers do. Therefore, derivations that are frequent and regular to native speakers may not be so

to ESL learners. For instance, knowing the word “hope” may not necessarily mean that an ESL learner would know the word “hopeful”. Similarly, knowing the word “develop” may not mean that he knows “-ment” could be added to it to create the noun form “development”. In this sense, lemmatising derivational suffixes may make ESL learners’ vocabulary size underestimated. Owing to this consideration, derivations were not lemmatised, and they were put into computer as valid data for scoring. The program to score Lex30 can be found at the website of Lognostics (2017).

4.3.3.5 Scoring of Writing Tasks

The computer program adopted in this research to score testees’ compositions was RANGE. Before the compositions were scored by RANGE, several steps were taken to process the data. Firstly, although the testees were asked to write a composition of no less than 200 words, only the first 200 words from the main body of the text were taken out for analysis. This was based on the finding of Laufer and Nation (1995) that the LFP over 200 words tends to be stable, whereas that of less than 200 words does not. All the compositions were transcribed into Microsoft Word File, and like in CAVLT and Lex30, spelling errors that did not affect word recognition were corrected. Then words in every composition were counted by the Microsoft Word File. The first 200 words (not including the title) were taken out for further treatment. Due to the fact that a few testees’ compositions were less than 200 words, these compositions were removed from valid data. As a result, 16 testees’ compositions, which were 32 pieces in total, were left for final analysis. Secondly, pseudo words were deleted, so that RANGE would not count them as words of “not in the lists” and bring about wrong results. Words that were semantically wrong were deleted as well.

It was due to the consideration that if a word is not used correctly in meaning, it should not be considered as known by the testee (Laufer & Nation, 1995). Additionally, to be consistent with the scoring standard of Lex30, proper nouns (e.g. Australia) were deleted, and they would not obtain points. This is consistent with the view of Laufer and Paribakht (1998) that proper nouns should not be regarded as a part of learners' vocabulary knowledge. Wrong derivations (e.g. "inhappy" for "unhappy") were deleted. It should be noted that Microsoft Word File only counts words that are correctly spelled. After the treatment, there were on average 201.4118 (SD=1.46026) tokens for each composition left for analysis in the first month, and 202.4375 (SD=1.63172) tokens in the twelfth month. Then RANGE was run to conduct LFP calculation for each of the 32 compositions.

RANGE produces results in terms of tokens, types, and word families. The numbers and percentages of tokens may not be reliable for analysis, as tokens not only include the inflections and derivations, but also include all the repetitions of them. For example, if the word "runs" occurs three times in a text, then it is counted as three tokens. Therefore, the numbers and percentages of tokens at different frequency levels may make researchers draw unreliable conclusions on FAV's frequency distributions. The numbers and percentages of types and word families may not be suitable to be analysed directly either. With regard to types, the base forms (e.g. develop), inflections (e.g. developing) and derivations (e.g. development) are all considered to be different types. If inflections are regarded as different words from their base forms, the FAV size will be overestimated. Therefore, further treatment was conducted on the data to remove inflections as individual types. Based on the list of types presented in the results of RANGE, all inflections were transferred into base forms. Then together with their base forms, they were counted as 1 type. For example, if the list of

types produced by RANGE showed that in the text there were types of “run”, “runs” and “running”, then “runs” and “running” were lemmatized to be “run”, and together the three types were regarded to be one type. Similarly, if a text had types of “develops” and “developing”, they were transferred to be “develop” and counted to be one type. The list of types of each composition was observed in this way, and extra types were manually deleted. Therefore, in this sub-study, the numbers and percentages of “types” refer to those after the treatment, not the original numbers and percentages of types calculated by RANGE.

The results of word families presented by RANGE need some manual treatment as well. As words of “not in the lists” do not belong to any of the three lists available in RANGE, the program is unable to calculate the numbers and percentages of word families that fall out of the 3 available word lists. Therefore, based on the types of “not in the lists” presented by RANGE, word families were counted by the student investigator. This manual processing was feasible, as there were a very limited number of types that belonged to “not in the lists” in each composition. This could be evidenced by the mean number of types being 5.3889 (SD=2.89297) in the first month, and the mean number of types being 5.5625 (SD=2.82769) in the twelfth month. Word families of “not in the lists” were counted, and then percentages of word families belonging to each of the 3 word lists in RANGE and “not in the lists” were calculated.

It is necessary to note that PVLТ, CAVLT, Lex30 and LFP in writing tasks all measure testees’ breadth of vocabulary rather than the depth of vocabulary. This is due to the fact that in these tests grammatical error is not a factor in scoring. It should also be noted that both Lex30 and LFP can only indicate FAV size indirectly. As is pointed out by Laufer (1998), it seems unlikely to devise a test that is able to measure

how many words a person can produce, unless his FAV repertoire is very small. However, the limited function of Lex30 and LFP is enough for this study, as the purpose of the sub-study is to investigate whether FAV grows in a 12-month period of time, not to calculate exactly how many FAV words the testees have.

4.3.4 Data Analysis

Statistical Package for Social Science (SPSS) of version 20.0 was adopted to conduct data analysis. All the scores of PVLTL, CAVLT, Lex30, and LFP were entered into SPSS for statistical processing. The methods adopted in this sub-study included Paired-sample T test and Correlation Analysis. Paired-sample T test was used to compare passive vocabulary size, controlled active vocabulary size, and free active vocabulary size between the first-month tests and the twelfth-month tests. It should be noted that in LFP analysis, the number of types and word families obtained from writing are frequency data, and Chi-square is usually adopted to compare frequencies. However, Chi-square requires that the Expected N be more than 5 (Theanalysisfactor, 2017). As in this research the relevant data could not meet this condition, Paired-sample T test was applied instead. In order to make T test appropriate to be used, frequencies were transferred into scores, in which one frequency was equal to one point. In addition to Paired-sample T test, Correlation Analysis was applied to investigate the relationship of passive vocabulary, controlled active vocabulary, and free active vocabulary.

4.4 Results

4.4.1 The Developmental Process of FAV in the 12-month Period of Time

The scores of Lex30 and LFP of types between the first month and the twelfth month were compared by Paired-sample T test. The results are demonstrated in Table 4.2.

Table 4.2. *Comparison of Lex30 and LFP in Terms of Word Types between the First Month and the Twelfth Month*

Test scores	Time to conduct the test	Number of subjects	Mean score	Standard deviation	t	df	P
Lex 30	The first month	21	82.4762	9.88240	0.614	20	0.546
	The twelfth month	21	81.2857	10.03067			
Word types of the first 1,000 frequency level	The first month	16	92.7500	6.53707	3.306	15	0.005
	The twelfth month	16	84.8750	7.58837			
Word types of the second 1,000 frequency level	The first month	16	6.8750	2.70493	-0.075	15	0.941
	The twelfth month	16	6.9375	2.69490			
Word types of the Academic Word List	The first month	16	7.8125	3.74555	-1.544	15	0.143
	The twelfth month	16	9.6875	5.26268			
Word types of “not in the lists”	The first month	16	5.6250	3.51900	0.574	15	0.574
	The twelfth month	16	5.1250	2.24722			

Results of Paired-sample T Test indicate that there was no significant difference between the first month's Lex30 score and the twelfth month's Lex30 score ($t=0.614$, $df=20$, $p=0.546$, $p>0.05$). According to the scoring principle of Lex30, response words that do not belong to the first 1,000 frequent content words, together with high-frequency function words, proper nouns and numbers would obtain points. Therefore, the result indicates that testees' FAV beyond the first 1,000 frequent content words, high-frequency content words, proper nouns and numbers did not grow in the 12-month period of time.

LFP of types from writing tasks provides more information on FAV development. According to Table 4.2, word types in the first 1,000 frequency band decreased significantly in the 12-month period ($t=3.306$, $df=15$, $p=0.005$, $p<0.05$). However, there was no significant difference in words from the second 1,000 frequency band, academic words, and "not in the lists". This shows that testees' FAV in the second 1,000 frequency level, the academic level, and beyond did not increase in the 12-month period of time. The results of LFP in terms of word families are presented in Table 4.3.

Table 4.3. *Comparison of LFP in Terms of Word Families between the First Month and the Twelfth Month*

Test scores	Time to conduct the test	Number of subjects	Mean score	Standard deviation	t	df	P
Word families of the first 1,000 frequency level	The first month	16	82.8750	4.61700	3.088	15	0.007
	The twelfth month	16	77.4375	6.69297			
Word families of the second 1,000 frequency	The first month	16	6.8125	2.71339	0.000	15	1.000
	The twelfth month	16	6.8125	2.76209			

level							
Word families of the Academic Word List	The first month	16	7.6250	3.51900	-1.316	15	0.208
	The twelfth month	16	9.2500	5.25991			
Word families of “not in the lists”	The first month	16	5.3125	3.40037	0.668	15	0.514
	The twelfth month	16	4.7500	2.08167			

In terms of percentages, in the first month the average percentage of types belonging to the first 1,000 frequency level was 82.4519%, to the second 1,000 frequency level was 5.0156%, to the academic words was 6.7806%, and to “not in the lists” was 4.8525%. In the twelfth month, the average percentage of types belonging to the first 1,000 frequency level was 80.5750%, to the second level was 6.3025%, to the academic words was 8.5819%, and to “not in the lists” was 4.5375%. In word families, the LFP distribution was similar. In the first month, the average percentage of word families in the first 1,000 frequency band was 81.0058% (SD=5.56919), in the second 1,000 frequency band was 6.6040% (SD=2.49476), in academic words was 7.3318% (SD=3.20586), and in “not in the lists” was 5.0583% (SD=3.01265). In the twelfth month, the average percentage of word families of the first 1,000 frequency band was 79.1893% (SD=6.2427), of the second 1,000 frequency band was 6.8991% (SD=2.49682), of the academic words was 9.1464% (SD=4.63597), of “not in the lists” was 4.7652% (SD=2.10334). No matter in terms of types or families, the average percentages of words belonging to the first 1,000 frequency level were around 80% in both the first-month test and the twelfth-month test. By contrast, the average percentage of words used by native speakers is around 70% (Cobb, 2003). It is indicated that even when L2 learners reach the upper-intermediate to advanced

proficiency level, they still heavily depend on the very high frequency words to use.

In addition, there is still a big gap between native speakers' FAV size and L2 learners' FAV size.

4.4.2 The Correlation of FAV with Passive Vocabulary and Controlled Active Vocabulary

In order to investigate the correlation of FAV with PV and CAV, the developmental process of PV and CAV was explored first (Table 4.4 and Table 4.5).

Table 4.4. *Comparison of PVL T Scores between the First Month and the Twelfth*

Month

The test score	Time to conduct the test	Number of subjects	Mean score	Standard deviation	t	df	P
Total score of PVL T	The first month	21	108.3333	10.42273	-4.811	20	0.000
	The twelfth month	21	116.9048	8.73444			
Score of Section A in PVL T (the 2,000 frequency level)	The first month	21	29.0952	0.99523	-1.482	20	0.154
	The twelfth month	21	29.5238	0.81358			
Score of Section B in PVL T (the 3,000 frequency level)	The first month	21	26.5714	1.98926	-3.325	20	0.003
	The twelfth month	21	27.5714	1.50238			
Score of Section C in PVL T (the 5,000 frequency level)	The first month	21	18.8095	4.74994	-7.469	20	0.000
	The twelfth month	21	23.0952	3.19225			
Score of Section D in	The first month	21	6.1429	5.46155	-1.356	20	0.190

PVLT (the 10,000 frequency level)	The twelfth month	21	7.8095	4.51242			
Score of Section E in PVLT (the academic words)	The first month	21	27.7143	2.32686	-2.449	20	0.024
	The twelfth month	21	28.9048	1.84132			

In Table 4.4, it can be seen that the receptive vocabulary grew significantly in the 12 months ($t=-4.811$, $df=20$, $p=0.000$, $p<0.05$). With regard to different sections, only scores in Section A and Section D did not increase significantly. Scores in other sections, including Section B, Section C, and Section E, got significant growth. In other words, receptive words at 2,000 and 10,000 frequency level did not get improvement, whereas receptive words at 3,000 level, 5,000 level, and academic level improved significantly.

Table 4.5. *Comparison of CAVLT Scores between the First and the Twelfth Month*

The test score	Time to conduct the test	Number of subjects	Mean score	Standard deviation	t	df	P
Total score of CAVLT	The first month	21	230.4762	31.04290	-7.688	20	0.000
	The twelfth month	21	268.0000	33.04694			
Score of Section A in CAVLT (the 2,000 frequency level)	The first month	21	77.4762	9.92280	-2.528	20	0.020
	The twelfth month	21	82.6190	9.81568			
Score of Section B in CAVLT (the 3,000 frequency level)	The first month	21	62.3333	7.46548	-2.214	20	0.039
	The twelfth month	21	69.1429	13.37268			

Score of Section C in CAVLT (the 5,000 frequency level)	The first month	21	36.0000	13.53514	-4.492	20	0.000
	The twelfth month	21	47.1429	10.44646			
Score of Section D in CAVLT (the university words)	The first month	21	51.4286	9.61026	-2.138	20	0.045
	The twelfth month	21	55.9524	6.84453			
Score of Section E in CAVLT (the 10,000 frequency level)	The first month	21	3.2381	5.18560	-4.551	20	0.000
	The twelfth month	21	13.1429	12.04278			

Table 4.5 compares scores of CAVLT between the first month and the twelfth month. The table shows that CAVLT scores increased significantly in the 12-month period ($t=-7.688$, $df=20$, $p=0.000$, $p<0.05$). With regard to scores in each section, all the sections' scores showed significant growth, including Section A ($t=-2.528$, $df=20$, $p=0.020$, $p<0.05$), Section B ($t=-2.214$, $df=20$, $p=0.039$, $p<0.05$), Section C ($t=-4.492$, $df=20$, $p=0.000$, $p<0.05$), Section D ($t=-2.138$, $df=20$, $p=0.045$, $p<0.05$) and Section E ($t=-4.511$, $df=20$, $p=0.000$, $p<0.05$). It is indicative that testees' controlled active vocabulary at 2,000 frequency level, 3,000 frequency level, 5,000 frequency level, 10,000 frequency level, and academic level all grew significantly in the 12-month period of time.

From Table 4.4 and Table 4.5, it can be observed that the scores of PVLT and CAVLT decreased with the decrease of words' frequency levels. In the first-month PVLT, the order of the mean scores in different sections was Section A (the 2,000 level, $mean=29.0952$, $SD=0.9952$)>Section B (the 3,000 level, $mean=26.5714$, $SD=1.98926$)>Section C (the 5,000 level, $mean=18.8095$, $SD=4.74994$)>Section D

(the 10,000 level, mean=6.1429, SD=5.46155); in the twelfth-month PVLТ, the order was Section A (the 2,000 level, mean=29.5238, SD=0.81358)>Section B (the 3,000 level, mean=27.5714, SD=1.50238)>Section C (the 5,000 level, mean=23.0952, SD=3.19225)>Section D (the 10,000 level, mean=7.8095, SD=4.51242). CAVLT in the first and twelfth month showed the same tendency. In the first-month CAVLT, the order of mean scores in different sections was Section A (the 2,000 level, mean=77.4762, SD=9.92280)>Section B (the 3,000 level, mean=62.3333, SD=7.46548)>Section C (the 5000 level, mean=36.0000, SD=13.53514)>Section E (the 10,000 level, mean=3.2381, SD=5.18560); in the twelfth-month CAVLT, the order was Section A (the 2,000 level, mean=82.6190, SD=9.81568)>Section B (the 3,000 level, mean=69.1429, SD=13.37268)>Section C (the 5,000 level, mean=47.1429, SD=10.44646)>Section E (the 10,000 level, mean=13.1429, SD=12.04278). This result is consistent with the view of Schmitt et al. (2001) that words are grasped in sequence, i.e., more frequent words tend to be learned earlier, and less frequent words tend to be learned later.

Pearson Correlation was calculated for scores of Lex30, PVLТ, and CAVLT. Results show that in the first month, Lex30 and PVLТ were not significantly correlated ($r=0.374$, $p=0.095$, $p>0.05$), while Lex30 and CAVLT were significantly correlated at a median level ($r=0.531$, $p<0.013$, $p<0.05$). Meanwhile, PVLТ and CAVLT were significantly correlated, and the correlation reached median level ($r=0.512$, $p=0.018$, $p<0.05$). With regard to the correlation of Lex30 with different sections in PVLТ and CAVLT, it turned out that Lex30 was only significantly correlated with Section B (the 3,000 frequency level) in CAVLT ($r=0.451$, $p=0.040$, $p<0.05$), but not correlated with other sections in PVLТ and CAVLT. From this result, it could be concluded that Lex30 in the first month was generally not correlated with PVLТ and CAVLT (except

for Section B in CAVLT), but PVLТ and CAVLT were significantly correlated with each other. In the twelfth month, Lex30 was not significantly correlated with either PVLТ ($r=0.013$, $p=0.9567$, $p>0.05$), or CAVLT ($r=0.057$, $p=0.807$, $p>0.05$). However, PVLТ and CAVLT were significantly correlated, and the correlation reached large level ($r=0.708$, $p=0.000$, $p<0.05$). No scores of individual sections in PVLТ and CAVLT was correlated with scores of Lex30.

In sum, in this sub-study the testees' passive vocabulary and controlled active vocabulary kept growing, but free active vocabulary experienced stagnation in the 12-month period of time; the testees' passive vocabulary and controlled active vocabulary were highly correlated, but they were not correlated with free active vocabulary. The finding suggests that being able to recognise and spell a word does not necessarily mean being able to actively produce the word.

4.5 Discussion

4.5.1 The Developmental Process of FAV

A large amount of information on FAV's developmental process is obtained from this sub-study. Firstly, it turns out that the Lex30 score did not change significantly in the 12-month period of time. According to the scoring standard of Lex30, only words beyond the first 1,000 frequent content words, high-frequency function words, proper nouns and numbers would get points. This means that in the 12 months, words beyond these categories did not grow significantly. Secondly, results by RANGE showed that in writing, word types at the first 1,000 frequency level decreased significantly, but word types at the second 1,000 frequency level, academic words and beyond did not increase significantly. It seems to be a good tendency that the testees

tended to use less words belonging to the first 1,000 frequency band after 12 months. However, the whole picture is not as optimistic, as words beyond this level failed to have satisfactory growth. This result indicates that the improvement of the testees' repertoire of words in the second 1,000 frequency level and beyond is limited. The result on word families shows the same tendency: word families of the first 1,000 frequency level decreased significantly, whereas word families of the second 1,000 frequency level, academic words, and beyond did not increase significantly. It is indicative that although the testees tended to use less words from the first 1,000 most frequent words, they failed to use more words from other frequencies after 12 months. If a text comprises primarily the most basic vocabulary, the expression in writing will be limited and not authentic enough.

Some findings in this sub-study can be compared with those in previous research. Firstly, different from this sub-study, Laufer (1991) found that participants' academic words got significant growth. The participants in Laufer (1991)'s study were year-one university students from English department. According to Laufer (1991), they were required to have a large amount of writing on language and literature in the first year. In this sub-study, the participants were the year-three to year-four university students from English major. In the year of the sub-study, the participants were required to have some writing practice as well. For example, they needed to complete a piece of academic thesis of around 5,000 words for graduation. The thesis was under a teacher's supervision, and the whole process included revisions for at least three times according to the supervisor's feedback. In addition, they had a compulsory unit of writing, which took 2 hours each week and lasted for one semester. In this unit writing practice was conducted. However, even with these writing practices, the academic words did not improve. The different results of the two studies may be

primarily owing to the fact that the participants in the two studies were at different learning stages. The input and output activities may promote academic words' growth more easily at low and intermediate proficiency levels than at upper-intermediate to advanced levels, as the abundant input and output seem not to work as efficiently as before. Secondly, this result is consistent with that of Laufer (1998). In Laufer (1998)'s research, participants' FAV seemed to "fossilize" even when their passive vocabulary and controlled active vocabulary increased.

In sum, the results of Lex30 test and LFP in this sub-study provide a relatively comprehensive picture on FAV's developmental process in a 12-month period of time. During this period, words (no matter in terms of types or families) at various levels seemed to experience stagnation.

4.5.2 The Relationship of FAV with Passive Vocabulary and Controlled Active Vocabulary

To investigate the relationship between different types of vocabulary knowledge, it is necessary to investigate the developmental process of receptive vocabulary and controlled active vocabulary at first. In the 12-month period of time, the testees' passive vocabulary and controlled active vocabulary grew significantly. With regard to passive vocabulary, except for words in the 2,000 level and in the 10,000 level, words of all other levels increased significantly. These included the 3,000 level, the 5,000 level, and the academic words. The reason why words from the 2,000 frequency level did not get significant growth may be that the high-frequency words had been grasped well to upper-intermediate to advanced ESL learners. This can be evidenced by the very high mean score of Section A in the first-month PVLIT test (mean=29.095, SD=0.99523), and that in the twelfth-month PVLIT test (mean=29.523, SD=0.81358).

The total score of each section in PVLТ was 30 points. Therefore, the mean scores of Section A show that most of words from this level had been grasped receptively, and there was almost no more space for further development. The lack of growth for words in the 10,000 frequency level may be due to the fact that this is a very low frequency level, and learners may not have many chances to encounter these words in their study. Therefore, the score in the 10,000 level did not change significantly, and it maintained low. It can be evidenced by the mean score in the first-month PVLТ test (mean=6.1429, SD=5.46155) and that in the twelfth-month test (mean=7.8095, SD=4.51242). The increase of words at the 3,000 and 5,000 level in PVLТ indicates that the testees' capability to read authentic English texts had improved (Schmitt et al., 2001); the increase of academic words in PVLТ test shows that their ability to read English academic text improved after 12 months as well. With regard to the controlled active vocabulary, words at all frequency levels developed significantly. This indicates that in the 12 months of time, the testees not only could recognise more words, but also could spell more words. Additionally, the Pearson Correlation analysis shows that testees' PV and CAV were significantly correlated with each other in both the first-month test and the twelfth-month test. However, FAV seems not to have a close relationship with receptive vocabulary and controlled active vocabulary, as it turns out that FAV was not significantly correlated with receptive vocabulary in the first and twelfth month, and with controlled active vocabulary in the twelfth month. Although FAV was correlated with controlled active vocabulary in the first month, further analysis shows that it was only correlated with Section C, not with any other sections in CAVLT. In other words, in the first month spelling seemed to be correlated with vocabulary production, but this correlation was rather limited.

The sub-study thus answers the second proposed question as well. FAV is not correlated with passive vocabulary, and FAV is generally not correlated with controlled active vocabulary either. However, passive vocabulary and controlled active vocabulary seem to be closely correlated with each other. In addition, both passive and controlled active vocabulary grew significantly in the 12 months. It is indicative that at least in upper-intermediate to advanced proficiency stage, the development of ESL learners' recognition and spelling capability does not promote their FAV growth. This finding is consistent with Laufer (1998)'s finding that FAV does not grow with the growth of passive and controlled active vocabulary. In Laufer (1998), two groups of subjects participated in the study. One group was consisted of 26 16-year-old 10th graders in Israel, who had been learning English for six years. The other group were 22 17-year-old students in the 11th grade in Israel, who had been studying English for seven years. Just like the Chinese subjects in this sub-study, English was a foreign language in Israel, and there was no English environment in their everyday life. These participants primarily obtained the English input from formal classes. By comparing FAV beyond the first 2,000 frequency level, Laufer (1998) found that there was no significant difference between the two groups in FAV size. And Laufer (1998) discovered that the 11th graders' passive vocabulary and controlled active vocabulary were significantly higher than those of 10th graders. Based on findings in Laufer (1998) and this research, it may be concluded that the FAV threshold tends to occur periodically, whereas passive vocabulary and controlled active vocabulary tend to develop in a more linear way.

However, the finding seems to be contrary to Meara and Fitzpatrick (2000)'s finding that the larger the receptive vocabulary is, the more the FAV will be. Taken a closer observation, it may be seen that the reliability of Meara and Fitzpatrick (2000)'s study

seems to be questionable. In the study, the variable of L2 proficiency may not be well controlled. The participants were 46 adult learners at different proficiency levels. The proficiency level seems to be an important factor affecting results of vocabulary tests. Therefore, the conclusion of Meara and Fitzpatrick (2000) may not be convincing as the variable of proficiency is not controlled.

This sub-study not only proves the Threshold Hypothesis proposed by Laufer (1991), but also obtains more information on the FAV threshold phenomenon. Firstly, the threshold seems to be long-lasting, since in the 12 long months of the sub-study FAV did not grow. Secondly, the stagnation of FAV seems to be more serious than our expectations: it is widespread, almost occurring at all frequency levels, and even at the frequency level as high as the second 1,000 frequency band. Thirdly, the FAV threshold phenomenon seems to be “stubborn”, as in the 12 months passive and controlled active vocabulary kept growing, but FAV was not promoted by the development of the two. Fourthly, combined with results of other research, it can be concluded that FAV threshold may occur periodically and at various proficiency stages.

These features of FAV threshold phenomenon provide some pedagogical implications to ESL learners. Firstly, teachers should be conscious that the FAV threshold phenomenon may occur, especially when ESL learners reach upper-intermediate to advanced proficiency level. Knowing the existence of the threshold may be the first step to tackle the problem. Secondly, teachers should be conscious that the threshold may be a serious problem in ESL learning: it is long-lasting, it occurs widely, and it occurs at most frequency levels. Thirdly, the efficiency of traditional teaching and learning methods to tackle the problem seems to be limited. Although it may easily

lead people to think that strengthening word recognition and spelling will help promote FAV development, the results of this sub-study indicate that at least when students reach upper-intermediate to advanced level of L2 proficiency, it is not the case. In other words, teachers should keep in mind that promoting FAV growth by strengthening passive and controlled active vocabulary may not be an efficient effort.

4.6 Conclusion

This chapter has reported the findings of a 12-month longitudinal sub-study to explore the developmental process of FAV, and its relationship with PV and CAV. The findings have answered the proposed research questions. It was found that among a group of upper-intermediate to advanced ESL learners, FAV stagnated in the 12-month period of time. It was not correlated with either PV or CAV. Both PV and CAV kept growing in the 12-month period, but FAV did not. Several features were detected in the FAV developmental process. Firstly, the stagnation of FAV seems to be long-lasting, since the sub-study lasted for 12 months and found no growth in FAV; secondly, the stagnation of FAV seems to be “stubborn”, as it not only lasted for a long time, but also was not related with the growth of other types of vocabulary. Thirdly, the FAV threshold tends to occur widely, as it happened at most frequency levels. This chapter empirically supports the Active Vocabulary Threshold Hypothesis raised by Laufer (1991), and obtains more information on the threshold phenomenon. In the following chapters, more quantitative and qualitative research will be carried out to investigate factors that may cause the threshold phenomenon.

Chapter 5

Comparison of Lexical Network between Frequently and Seldom Produced Words

5.1 Introduction

As it has been found in Chapter 4, not all receptive words could smoothly develop into productive words; some of them stagnate in the process, and fail to become productive. In order to explore the reasons for the stagnation, it is necessary to examine the mental representation of receptive and productive words, and investigate whether there are differences between the two. Unfortunately, it is difficult to pick receptive and productive words from learners' mental lexicon. This is because the mental lexicon is like a dark box and cannot be examined directly. Even if a word is never used by a person in the past, researchers should be cautious to conclude that the word is not a productive word. There is still possibility that the word will be retrieved

and produced in a specific circumstance. The alternative idea of this sub-study is that although receptive and productive words cannot be directly picked out, words that are frequently produced by L2 learners, and words that are seldom or almost never used by them may be picked in some way. If a word is a high-frequency word in authentic English texts and have no cultural strangeness, it can be postulated that the word is frequently needed in similar L2 contexts as well. Therefore, if both a group of frequently produced words (FPW) and seldom produced words (SPW) are high-frequency words in authentic English texts, and if learners are familiar with both of them, the reason why one group are used frequently by the learners and why the other group are not is an interesting issue. The difference between the two groups of words in mental representation may shed some light on the lexical network construction procedure in FAV development.

This chapter attempts to investigate the issue posed above. It first raises research questions, then proposes research methods, including the research instruments, data collection procedures, and data analysis methods. At the end of the chapter, results of the sub-study are reported and discussed, and conclusions are presented.

5.2 Research Questions

The primary question in this chapter is: What are the differences (if any) of the mental lexicon organisation between frequently produced words and seldom produced words? In order to be more manageable, the general question is broken into three sub-questions:

- (1) Do semantic nodes and non-semantic nodes distribute differently in FPW and SPW? What are the differences (if any)?

- (2) Do different types of semantic nodes distribute differently in FPW and SPW? What are the differences (if any)?
- (3) Do different types of non-semantic nodes distribute differently in FPW and SPW? What are the differences (if any)?

5.3 Research Method

This section presents major components of the research method adopted in this sub-study, including instruments, procedures of data collection, data preparation, and data analysis. The background information for participants participating in this sub-study has been provided in Chapter 3.

5.3.1 Instruments

Four instruments were adopted in this sub-study. The instruments included two Vocabulary Knowledge Scale (VKS) tests and two word association tests. These tests are introduced in this section.

5.3.1.1 Vocabulary Knowledge Scale Test

It has been repeatedly found that familiarity of words has a significant effect on the response pattern (Entwisle, 1966; Namei, 2004; Wolter, 2001; Zareva & Brent, 2012). To control this extraneous variable, this sub-research adopted Wesche and Paribakht's (1996) Vocabulary Knowledge Scale (VKS) test to measure testees' familiarity level to candidate stimulus words in word association tests. In the test, five scales were given after a target word was presented:

- I. I don't remember having seen this word before;
- II. I have seen this word before but I don't know what it means;
- III. I have seen this word before and I think it means ____; (synonym or translation)

IV. I know this word. It means ____; (synonym or translation)

V. I can use this word in a sentence, e.g.: _____. (write a sentence)

Scores corresponding to choice I, II, III, IV and V were 0, 1, 2, 3, and 4 respectively. However, as has been pointed out in Section 2.3.2.2, the testee's choice of the scale did not necessarily mean that they would get the corresponding score; it also depended on what they filled in the following blanks. For example, if the testee chose "III" but gave a wrong synonym or translation in the following blank, he would not get a score of 2, but 1 instead; if the testee chose "IV" but gave a wrong synonym or translation in the following blank, he would not get a score of 3, but only 1. In sum, it was not only the scale testees chose, but also their answer to the blank would decide the score they could get. The VKS tests and the word association tests used in this sub-research are presented in Appendix F and Appendix G.

It should be noted that with the tested word being provided, the familiarity level measured by VKS test only covers recognition and controlled production of the tested word. As the target word has been given in the test, free active production of the word is not tested. In other words, the word tested to have high familiarity level by VKS test can possibly be FPW, or SPW.

5.3.1.2 Word Association Test

Based on the research aim of this sub-study, which was to compare lexical organisations of SPW and FPW, this sub-study adopted word association tests as the research instrument. In word association tests, there are a number of factors affecting testees' reaction pattern. These factors include whether the stimulus words are cognates in testees' native language (Fitzpatrick & Izura, 2011; Wolter, 2001), whether they tend to trigger set responses (Fitzpatrick et al., 2015), stimulus words'

lexical class (Clark, 1995; de Groot, 1992; Deese, 1962; Fitzpatrick et al., 2015; Källkvist, 1999; Lotto & de Groot, 1998; Nissen & Henriksen, 2006; Sökmen, 1993), cultural strangeness (Carolyn et al., 2004; Chow et al., 1987; Son et al., 2014), concreteness (de Groot, 1989, 1993; Fitzpatrick et al., 2015; Kolers, 1963; D. L. Nelson & Schreiber, 1992; van Hell & de Groot, 1998), and familiarity (Entwisle, 1966; Namei, 2004; Wolter, 2001; Zareva & Brent, 2012). Therefore, in order to ensure the validity and reliability of the sub-research, these variables need to be effectively controlled.

To achieve this goal, several measures have been taken. Firstly, the General Service List (GSL) was referenced from the website of Lextutor (2017). From its second 1,000 most frequent word band, pairs of words containing similar meanings were picked out. The exclusion of the first 1,000 band in word selection was due to the consideration that words from that frequency level tend to elicit dominant responses, which may conceal the characteristics of lexical representations (Wolter, 2001). Limiting the selection to the second 1,000 most frequent words ensured that participants could recognise, and in most cases were familiar, with the cues (Wolter, 2001). In addition, having the same frequency meant the participants had similar chances to encounter these words in English texts in reading and listening.

The selection criteria were: one word in the pair is frequently used by ESL students on college level, whereas the other word in the pair, even though having similar meanings and from the same frequency band, is seldom or never used. In this step, 13 pairs of words were selected. Due to the fact that these pairs of words shared similar meanings and were chosen from the same band, it could be ascertained that they were of the similar frequency, concreteness, and most likely, they were of the same primary lexical class. However, it should be noted that in this stage the word selection was

based on the researcher's subjective judgement. Although the researcher in this study has had a number of years' ESL teaching experience at universities, the judgement was subjective, and objective standards were needed.

Therefore, in the next step, the *Spoken and Written English Corpus of Chinese Learners* (2.0) by Wen, Wang, and Liang (2009) was referenced. The corpus collected 6,579 spoken and written compositions done by Chinese university students, from both English and other majors. Each word from the 13 selected pairs was searched in the corpus with the use of the software AntConc3.3.5w. The software counted the target words' frequency of occurrences in the corpus. Based on the results, two pairs of words were removed, and 11 pairs were kept for further selection. All the remaining 11 pairs shared the feature that one word in each pair appeared for a number of times (496 times on average) in the corpus, whereas the other one almost never appeared (0 time on average). This indicated that one word from each selected word pair was frequently used by Chinese university students, but the other word was not. This step was more objective than ESL teachers' subjective judgement by teaching experience. It further ensured that the selected words were suitable for the sub-study.

Thirdly, a pilot study was conducted. 11 students in the university, who were in the same year level and from the same major as the formal participants, participated in this sub-study. In the pilot study, Wesche and Paribakht's (1996) VKS test was administered. According to the results, words to which some students showed low familiarity were excluded. Words having more than one frequently used meanings were removed as well. For example, it was found that the word "bold" had two frequently used meanings, one was "brave", and the other was "darker and thicker for font". Therefore, the candidate word "bold" was excluded from the cues. The reason

to remove these words was to avoid confusions in response categorisation. After the selection, 5 pairs of words were left to be stimulus words. Due to the fact that the participants in the pilot study were in the same year level and in the same major of the university, they were considered to be at roughly the same English proficiency level, i.e. upper-intermediate to advanced level, as the formal testees in this sub-study were. Additionally, the selected words were all high-frequency words. Therefore, it was well based to assume that the formal testees were familiar with those cue words as well. However, even with these considerations, VKS tests would still be conducted in the formal sub-study. If some words turned out to have low familiarity, they would be removed from data analysis.

Although efforts had been made in FPW and SPW selection, other intervening variables still existed. It was possible that there were individuals who could not produce the words as others did, and it was possible that there were individuals who frequently used the words that were not produced by most of the L2 learners in similar proficiency levels. However, it was believed that this possibility was low. Firstly, the researcher had been teaching ESL in China for over ten years. Therefore, the researcher had rich experience on the students' interlanguage of English in China. Secondly, the corpus adopted in this sub-study collected more than 6,500 speaking and writing texts by Chinese L2 learners from public universities. These students shared with the participants the age, native language, and language environment in China. In addition, they all experienced standard elementary and middle school education, and then passed the national college entrance examination in China, which was a requirement for higher education. After entering college, the standard national syllabus was followed in college teaching. Therefore, the corpus can represent the general English usage patterns of L2 learners in universities in China.

The five pairs of selected words covered nouns, adjectives, verbs, and adverbials. Among them, there were three pairs of abstract words and two pairs of concrete words. The level of concreteness was judged by the concreteness scale of P. Zhang (2010a), which had been introduced in Chapter 2. None of the selected words was cognates of Chinese, neither were they culturally strange to the testees. None of them was very short words or very long words, i.e., words that were less than three letters long or words that were more than ten letters long. The key information on the selected stimulus words is demonstrated in Table 5.1 and Table 5.2.

Table 5.1. *Key Information on Selected FPW*

FPW	Word class	concreteness
afraid	adj.	abstract
damage	v.	abstract
happy	adj.	abstract
quickly	adv.	concrete
soil	n.	concrete

Table 5.2. *Key Information on Selected SPW*

SPW	Word class	concreteness
cowardly	adj.	abstract
wreck	v.	abstract
amused	adj.	abstract
hastily	adv.	concrete
clay	n.	concrete

5.3.2 Data Collection Procedures

5.3.2.1 Pilot Study

Before the formal data collection, a pilot study was carried out. In the pilot study, 11 year-three English-major students, who were studying in the same university as formal testees were asked to do the VKS test and word association tests. There were several purposes of the pilot study. The first purpose was to help researchers find suitable stimulus words for the formal sub-study. The second purpose was to ascertain

that the test requirements and test items were clear enough to the testees. The third purpose was to help researchers find the appropriate length of time for formal tests. According to the results of the pilot study, ten words were selected as cue words. Based on the students' feedback obtained in the pilot study, minor changes were made to make test requirements clearer to testees. First, to avoid misunderstandings and confusions, the test instructions were provided in Chinese. Secondly, more details were provided in the instruction. In the word association section, testees were told that there was no limit on the type of association, and they were asked to write down any association that came into their minds when they encountered the cue words. In the vocabulary knowledge scale section, testees were reminded to fill in the blanks if their choice was III, IV, or V. In addition, an example on how to fill in the blanks was provided. In both the word association section and the vocabulary knowledge scale section, testees were reminded not to use dictionaries when they did the tests. Based on the length of time spent in the pilot tests, it was decided that in the formal sub-study each VKS test would take two minutes, and each word association test would take one and a half minutes.

5.3.2.2 Formal Data Collection

The formal data collection was conducted in May and June, 2015. The reason to choose this time to collect data was due to the consideration that usually subjects were in the mid semester at that time; therefore, the results could reflect their normal performance. Before the data were collected, the student investigator orally explained the requirements for VKS test and word association test, and written instructions on those tests were presented to testees as well. Moreover, the testees were encouraged to ask questions on the test format and requirement until there was no doubts and

confusions. After ascertaining that all subjects had understood the test format and requirements correctly, the formal tests were conducted. Since the stimulus words in the SPW group and FPW group shared similar meanings, the tests were conducted at two different times to avoid reminding effects. The interval between the tests was one month, so that the first VKS and association test would have little influence on the second one. Meanwhile, the interval time was not long enough for the testees to have a significant change in their vocabulary size and L2 proficiency. The first VKS and word association test was for SPW, and the second was for FPW. Both tests were conducted out of class. To avoid audio or oral mistakes, word association tests adopted look-and-write pattern, in which the participants were asked to instantly write down three words that came to their mind as soon as they saw the prompt word. Guided by the pilot study, the student investigator told testees not to change the answers that they had already written down; if they could not make any association when encountering the prompt word, they should just leave the blank empty. In addition, they were told that there was no “right” or “wrong” answers to the prompts. Dictionaries were forbidden in both VKS and word association tests.

It should be noted that the “chain phenomenon”, in which testees’ response was not a reaction to the stimulus word, but rather a reaction to the former response that they gave in the test, appeared in some research such as P. Zhang (2011), and Nissen and Henriksen (2006). To avoid this phenomenon, some researchers (e.g. P. Zhang, 2011) chose to only analyse the first response. Some other researchers (e.g. Nissen & Henriksen, 2006) chose to consider chained response as a valid response in analysis, and classed it as non-semantic response. It may not be advocated in this sub-study, however, that a chained response be considered as valid for analysis. A chained response is a reaction elicited by the response before it, not a reaction to the prompt

word. It therefore offers little information on the prompt's lexical connection in mental lexicon. Moreover, in this sub-study the chain phenomenon only occurred occasionally. In FPW group, there were four chained responses, accounting for 1.33% of the total responses; in SPW group, there were five responses of this type, accounting for 1.67% of all responses. Therefore, the occurrence of chain phenomenon in this sub-study was rare and could be neglected. Moreover, observing three responses, instead of only one, may help researchers draw a more comprehensive picture of the word's lexical organisation (Schmitt, 1998). Taking all these factors into consideration, it was proposed in this sub-study that three responses to each stimulus word be observed and analysed; chained responses would be discarded as invalid data, and they would not be used in analysis.

5.3.3 Data Preparation

5.3.3.1 Selection of Valid Data

After the subjects finished all the tests, the test papers were collected. Then several steps were taken to select valid data. Firstly, the test papers were sorted into two groups, one was SPW group, and the other was FPW group. Each group contained a VKS and a word association test. Secondly, data done by the subjects who failed to participate in both times were removed. Only data by the subjects who had participated in both times, i.e., one time for SPW, and the other time for FPW, were kept. Thirdly, the test papers suspected to be done carelessly, such as those having a number of empty blanks, were discarded as well. After this selection procedure, 20 subjects' data were left for final analysis, resulting in 600 responses in word association tests.

5.3.3.2 Scoring the VKS Tests

The subjects' VKS tests were graded according to the scoring standard set up by Wesche and Paribakht (1996). If the subjects' choice was I or II, they would get a score of 0 or 1 respectively. If subjects' choice was III, IV, or V, their answer to the blank needed to be checked as well. If the answer to the blank was correct, a score of 2, 3, or 4 would be given to choice III, IV, and V respectively. However, if a subject chose III or IV, but gave a wrong synonym or translation for the target word, then he could only get 1 point; if he chose IV, but the sentence he composed was correct in word meaning but incorrect in grammar, he would get 2 points; if he chose IV, but wrote a sentence with wrong meaning for the target word, he could only get 1 point.

5.3.3.3 Scoring the Word Association Tests

Each response needed to be categorised before the test was scored. Categorisation was conducted according to the revised categorisation framework introduced in Chapter 2. After the classification, the frequency of occurrences for each type of response by every subject was counted. One occurrence would be given 1 point; therefore, the frequency of occurrences was equal to the points being obtained.

5.3.4 Data Analysis

The software used for statistical data analysis was Statistical Package for Social Science (SPSS), and the version of the software was 20.0. All scores of VKS and word association tests by the selected 20 subjects were entered into SPSS for statistical processing. Statistical processing methods in this sub-study included descriptive analysis, Paired-sample T test, and One-way Repeated Measures ANOVA. Descriptive Statistics were applied to portray a general picture of the subjects' familiarity with the target words, and to offer researchers descriptive

information on the subjects' response scores in word association tests. Paired-sample T test was adopted to compare familiarities between SPW and FPW group, and to compare response scores between the two groups. Paired-sample T test was used to compare scores of different response types within SPW and FPW group as well. In cases where more than two types of responses needed to be compared, One-way Repeated Measures ANOVA was adopted. It should be noted that usually Chi-Square is applied to compare frequencies between two or more groups. However, as has been pointed out in Chapter 4, when Chi-Square is used, it is expected that the sample size be above 150, since the value of Chi-Square is significantly affected by sample size; in addition, it is suggested that the Expected N should be more than 5 (Theanalysisfactor, 2017). As in this sub-research these two conditions cannot be met, Paired-sample T test and One-way Repeated Measures ANOVA were applied instead of Chi-square test. In order to use Paired-sample T test and One-way Repeated Measures ANOVA, the data were given further treatment. Frequencies of responses were transferred into scores, in which one occurrence was equal to one point. Since in word association tests subjects were asked to write down three responses to each of the five stimulus words, there was a total of 15 points for each subject.

5.4 Results

The results yielded by SPSS were reported in this section. The Paired-sample T test showed that there was no significant difference in the familiarity between SPW and FPW ($t=-1.000$, $df=19$, $p>0.05$). In addition, the mean familiarity scores of SPW and FPW were 24.250 ($SD=0.96655$) and 24.450 ($SD=0.99868$) respectively, indicating that subjects were familiar with both SPW and FPW.

5.4.1 Semantic and Non-semantic Distribution of SPW and FPW

To answer the first question posed in Section 5.2, mean scores of semantic and non-semantic responses were compared by Paired-sample T test, as displayed in Table 5.3 and Table 5.4.

Table 5.3. Comparison of Semantic and Non-semantic Responses in FPW and SPW

Word type	Response type	Number of subjects	Mean score	Standard deviation	t	df	P
FPW	Semantic	20	12.050	1.791	11.949	19	0.000
	Non-semantic	20	2.750	1.713			
SPW	Semantic	20	8.550	2.089	2.440	19	0.025
	Non-semantic	20	6.200	2.238			

Note: Each subject's full score was 15 points

Table 5.3 demonstrates that FPWs' semantic responses were significantly higher than their non-semantic responses ($p=0.000<0.05$). Meanwhile, semantic responses accounted for 81.4% of all responses in FPW group, and non-semantic responses accounted for 18.6%, indicating that FPW's lexical network was dominated by semantic links. With regard to SPW, semantic responses significantly outnumbered non-semantic responses as well ($p<0.05$). The percentage of semantic responses was 58%, and the percentage of non-semantic responses was 42%. This result indicated that although SPW's semantic connections outnumbered non-semantic connections, the advantage of semantic links in SPW was not as evident as that in FPW, as was revealed by the percentage of semantic and non-semantic responses. Past research has found that semantic links dominated in L1 mental lexicon. For example, Wolter

(2001) found that semantic links accounted for 92.7% in L1 mental lexicon, and P. Zhang (2010b) discovered that L1 testees' semantic responses reached 86.0% of all responses; Dong and Zhang (2011) examined responses in Edinburgh Association Thesaurus (EAT), which could be downloaded from the website of EAT (2015). In the thesaurus, each cue word was made associations by 100 native speakers. Dong and Zhang (2011) found that stimulus words' first most frequent responses were all semantic responses. Referencing these findings, it could be concluded that with regard to semantic and non-semantic distribution in lexical network, L2 FPW seem to be more similar to L1 words than L2 SPW do.

Table 5.4. *Comparison of Semantic and Non-semantic Responses between FPW and SPW*

Response type	Word type	Number of subjects	Mean score	Standard deviation	t	df	P
Semantic	FPW	20	12.05	1.791	-5.552	19	.000
	SPW	20	8.55	2.089			
Non-semantic	FPW	20	2.75	1.713	5.368	19	.000
	SPW	20	6.20	2.238			

Note: Each subject's full score was 15 points

As displayed in Table 5.4, the semantic responses of FPW were significantly higher than those of SPW ($p=0.000<0.05$); the non-semantic responses of FPW, on the other hand, were significantly lower than those of SPW ($p=0.000<0.05$).

5.4.2 Distribution of Different Types of Semantic Nodes in SPW and FPW

In order to explore the distribution of different types of semantic nodes in FPW and SPW, Paired-sample T test and One-way Repeated Measures ANOVA were adopted, as demonstrated in Table 5.5 and Table 5.6.

Table 5.5. *Descriptive Statistics for Different Types of Semantic Responses in FPW and SPW*

Word type	Response type	Number of subjects	Mean score	Standard deviation
FPW	Paradigmatic	20	8.900	1.071
	Syntagmatic	20	2.050	2.489
	Encyclopaedia	20	1.100	1.234
SPW	Paradigmatic	20	6.150	2.346
	Syntagmatic	20	0.800	0.768
	Encyclopaedia	20	1.600	1.789

Note: Each subject's full score was 15 points

An One-way Repeated Measures ANOVA was conducted to compare response scores of paradigmatic, syntagmatic, and encyclopaedia associations in both SPW and FPW. The means and standard deviations were presented in Table 5.5. As for FPW, there was a significant effect of response type (Wilks' Lambda=.104, $F [2,18]=77.366$, $p<0.01$, multivariate partial eta squared=.896). With regard to SPW, there was a significant effect of response type as well (Wilks' Lambda=.130, $F [2,18]=60.439$, $p<0.01$, multivariate partial eta squared=.870). This result indicated that the amount of paradigmatic, syntagmatic and encyclopaedia links were significantly different in both SPW and FPW.

In the FPW group, paradigmatic responses accounted for 73.86% of semantic responses, syntagmatic responses accounted for 17.01%, and encyclopaedia responses accounted for 9.13%. Ranking of the three types of responses from high to low was paradigmatic > syntagmatic > encyclopaedia. In the SPW group, paradigmatic responses accounted for 71.93% of semantic response, syntagmatic response was 9.36%, and encyclopaedia response was 18.71%. Ranking of the three types from high to low was paradigmatic > encyclopaedia > syntagmatic. This result was consistent with the findings of Zareva and Brent (2012) and Zareva (2007). Zareva and Brent (2012) and Zareva (2007) found that to advanced L2 learners or to words with high familiarity, paradigmatic connections would outnumber syntagmatic connections. This result also suggested that no matter to FPW or to SPW, paradigmatic responses would play a dominant role, accounting for more than 70% in semantic links; in addition, paradigmatic nodes' proportion from SPW to FPW was stable. Another finding was that the lexical structure of the two groups was different. In the FPW group, the testees showed a preference for paradigmatic and syntagmatic responses, whereas in the SPW group they showed a preference for paradigmatic and encyclopaedia responses. Therefore, from SPW to FPW the importance of syntagmatic nodes tended to elevate, whereas the importance of encyclopaedia nodes tended to decline.

Table 5.6. *Comparison of Different Types of Semantic Responses between FPW and SPW*

Response type	Word type	Number of subjects	Mean score	Standard deviation	t	df	P
Syntagmatic	FPW	20	2.050	1.234	-3.263	19	0.004
	SPW	20	0.800	0.768			

Paradigmatic	FPW	20	8.900	2.490	-3.832	19	0.001
	SPW	20	6.150	2.346			
Encyclopaedia	FPW	20	1.100	1.071	1.157	19	0.262
	SPW	20	1.600	1.789			

Note: Each subject's full score was 15 points

To compare the semantic organisation between the SPW and FPW, different types of semantic reactions, including syntagmatic, paradigmatic, and encyclopaedia responses were compared. Table 5.6 exhibits that paradigmatic responses in FPW group were significantly higher than those in the SPW group ($p=0.001<0.05$); syntagmatic responses in FPW group were significantly higher than those in SPW group as well ($p=0.004<0.05$). As for encyclopaedia responses, the two word groups did not show significant difference ($p=0.262>0.05$). Based on the statistics in Table 5.5 and Table 5.6, it could be concluded that from the SPW group to the FPW group paradigmatic and syntagmatic responses experienced significant increase, whereas the encyclopaedia responses did not; meanwhile, the proportion of syntagmatic responses in semantic connections tended to increase, whereas encyclopaedia responses tended to decrease. To have a closer observation, different types of paradigmatic and syntagmatic reactions were investigated. Table 5.7 provides results of Descriptive Statistics for different types of paradigmatic responses in FPW and SPW.

Table 5.7. *Descriptive Statistics for Different Types of Paradigmatic Responses in FPW and SPW*

Word type	Response type	Number of subjects	Mean score	Standard deviation
FPW	Synonym	20	6.750	2.268
	Antonym	20	0.700	0.801
	Hyponym	20	0.000	0.000
	Meronym	20	0.000	0.000
	Coordinate	20	0.850	0.988
	Ingredient	20	0.600	0.598
SPW	Synonym	20	3.800	1.673
	Antonym	20	0.200	0.523
	Hyponym	20	0.000	0.000
	Meronym	20	0.000	0.000
	Coordinate	20	1.150	0.988
	Ingredient	20	1.000	0.725

Note: Each subject's full score was 15 points

An One-way Repeated Measures ANOVA test was conducted to compare scores of different types of paradigmatic responses in both SPW and FPW. The means and standard deviations are presented in Table 5.7. With regard to FPW, there was a significant effect of response type (Wilks' Lambda=.063, $F[4,16]=59.561$, $p<0.01$, multivariate partial eta squared=.937). With regard to SPW, there was a significant effect of response type as well (Wilks' Lambda=.118, $F[4,16]=29.911$, $p<0.01$, multivariate partial eta squared=.882).

Table 5.7 exhibits that in FPW synonym had the highest rate of occurrence, accounting for 75.84% of all paradigmatic links; coordinate (9.55%) came the second, followed by antonym (7.87%) and ingredient (6.74%); the last two categories were hyponym and meronym, both had no occurrence at all. With regard to SPW, the order

was somewhat similar, with synonym (61.7%) being the largest category, followed by coordinate (18.70%), ingredient (16.26%) and antonym (3.25%); hyponym and meronym were the last two categories, accounting for 0% of all paradigmatic links. This result revealed the evident advantage of synonym links in both FPW and SPW, which supported the conclusion made by Liu et al. (2012) that synonym played an important role in L2 mental lexicon. Table 5.8 provides the comparison of different types of paradigmatic responses between FPW and SPW.

Table 5.8. *Comparison of Different Types of Paradigmatic Responses between FPW and SPW*

Response type	Word type	Number of subjects	Mean score	Standard deviation	t	df	P
Synonym	FPW	20	6.750	2.268	-4.914	19	0.000
	SPW	20	3.800	1.673			
Antonym	FPW	20	0.700	0.801	-2.236	19	0.038
	SPW	20	0.200	0.523			
Hyponym	FPW	20	0	0			
	SPW	20	0	0			
Meronym	FPW	20	0	0			
	SPW	20	0	0			
Coordinate	FPW	20	0.850	0.988	1.371	19	0.186
	SPW	20	1.150	0.988			
Ingredient	FPW	20	0.600	0.598	1.902	19	0.072
	SPW	20	1.000	0.725			

Note: Each subject's full score was 15 points

As it has been discussed in Chapter 2, paradigmatic category is further classed into hyponym (e.g. animal→cat; chair→furniture), meronym (e.g. hand→body), antonym (e.g. cold→hot), synonym (e.g. sad→upset), ingredient (e.g. china→clay), and

coordinate (e.g. cat→dog). Responses of these types between the FPW group and the SPW group were compared, which was demonstrated in Table 5.8. Results showed that there were no hyponym or meronym connections. Even the stimulus word “soil”, which may have more potential to trigger hyponym or meronym responses, did not elicit these two types of reaction. This may suggest that hyponym and meronym do not play significant roles in L2 lexical network construction. With regard to other types of paradigmatic connections, synonym ($p=0.000<0.05$) and antonym ($p=0.038<0.05$) in the FPW group were significantly higher than those in the SPW group.

The findings from Table 5.7 and Table 5.8 suggested that from SPW to FPW the number of synonym and antonym nodes tended to grow; the percentage of both synonym and antonym in paradigmatic responses tended to rise. However, they rose at different degrees, with synonym by 14.05%, and antonym by 4.62%; ingredient and coordinate nodes did not have significant increase in number, but both of them declined in percentage in paradigmatic responses, with ingredient declining by 9.52%, and coordinate by 9.15%. Therefore, different types of nodes seemed to change in different directions and at different degrees: some nodes tended to grow in percentage, while some tended to decline; some tended to increase in number, while some did not. The comparison of different types of syntagmatic responses in FPW and SPW is presented in Table 5.9.

Table 5.9. *Comparison of Different Types of Syntagmatic Responses in FPW and SPW*

Word type	Response type	Number of subjects	Mean score	Standard deviation	t	df	P
FPW	Compound/chunk	20	0.100	0.308	-5.663	19	0.000
	Syntactic	20	1.950	1.317			
SPW	Compound/chunk	20	0.000	0.000	-4.660	19	0.000
	Syntactic	20	0.800	0.768			

Table 5.9 suggests that different types of syntagmatic responses were significantly different in response amount in both SPW and FPW. Syntactic nodes were dominant in both FPW and SPW, accounting for 96.269% of syntagmatic connections in the SPW group, and 95.082% in the FPW group. Syntactic nodes are evidently important for word production, since words are usually produced in sentences, and syntactic principle is a critical factor to construct sentences (Tallerman, 2015). Moreover, when language users do not have adequate syntactic knowledge of the target word, they may consciously or unconsciously avoid using the word in order to prevent making mistakes. In terms of compound/chunk, the percentage of compound/chunk was small, with 4.918% in FPW, and 3.731% in SPW. The cue word “afraid” was selected to be observed individually, due to the consideration that “afraid” may have more potential to elicit compound/chunk response than other cues in this sub-study. This was because the set phrase “afraid of” was of high frequency in English usage. However, only 2 participants produced the response “of” to the cue “afraid”. It is believed by some researchers (e.g. Zareva & Brent, 2012) that compound/chunk links are inevitable parts in an efficient lexical network. However, results of this sub-study failed to confirm this view. It might result from the feature of the prompt words

adopted in this sub-study, as these words did not compose rich compounds or chunks with other words. Therefore, the issue may need further investigations. The comparison of different types of syntagmatic responses between FPW and SPW is presented in Table 5.10.

Table 5.10. *Comparison of Different Types of Syntagmatic Responses between FPW and SPW*

Response type	Word type	Number of subjects	Mean score	Standard deviation	t	df	P
Compound/chunk	FPW	20	0.100	0.308	-1.453	19	0.163
	SPW	20	0.000	0.000			
Syntactic	FPW	20	1.950	1.317	-2.881	19	0.010
	SPW	20	0.800	0.768			

Table 5.10 showed that the number of syntagmatic nodes in FPW were significantly higher than that of those in SPW ($p < 0.05$), but there was no significant difference in compound/chunk nodes between the two groups. The number of syntactic nodes had a significant increase in SPW-to-FPW process, but the number of compound/chunk nodes did not.

5.4.3 Distribution of Different Types of Non-Semantic Nodes in SPW and FPW

The comparison of different types of non-semantic responses between FPW and SPW is presented in table 5.11.

Table 5.11. *Comparison of Different Types of Non-semantic Responses between FPW and SPW*

Response type	Word type	Number of subjects	Mean score	Stand deviation	t	df	P
Phonological-formal	FPW	20	2.100	1.410	-5.080	19	0.000
	SPW	20	4.500	1.821			
Pseudo word	FPW	20	0.200	0.523	2.027	19	0.057
	SPW	20	0.600	0.821			
No response	FPW	20	0.100	0.308	2.483	19	0.023
	SPW	20	0.800	1.196			
Unclassifiable	FPW	20	0.350	1.137	-0.224	19	0.825
	SPW	20	0.300	0.470			
Translation	FPW	20	0.000	0.000			
	SPW	20	0.000	0.000			

Note: Each subject's full score was 15 points

Table 5.11 shows that there was no “translation” response in both the SPW and FPW groups. It seemed that the testees did not try to search for corresponding words in L1 mental lexicon when they encountered the cues, which may indicate that testees' L2 mental lexicon was not dependant on L1 mental lexicon. In addition, Table 5.11 shows that the “pseudo word” responses were sparse as well. The pseudo word links in the SPW group accounted for 4.07% of all responses, and in the FPW group accounted for 1.35%. The phonological-formal responses in the SPW group were significantly higher than those in the FPW group, and “no responses” in the SPW group were significantly higher than those in the FPW group. Other types of nodes did not have significant differences between the two groups. “No response” in the FPW group accounted for 0.68% of all responses, and in the SPW group accounted

for 5.42%, indicating that both FPW and SPW had set up rich nodes in lexical network. However, “no response” nodes in SPW were significantly higher than those in FPW, which implied that FPW’s lexical network was denser than SPW’s, and it had less empty points in the network. The number of pseudo words declined significantly, and its percentage in non-semantic connections declined significantly as well, suggesting that in SPW-to-FPW process the invalid wrong connections tended to decrease. The Descriptive Statistics for different types of non-semantic responses in FPW and SPW is provided in Table 5.12.

Table 5.12. Descriptive Statistics for Different Types of Non-semantic Responses in FPW and SPW

Word type	Response type	Number of subjects	Mean score	Standard deviation
FPW	No response	20	0.100	0.308
	Unclassifiable	20	0.350	1.137
	Phonological-formal	20	2.100	1.410
	Pseudo word	20	0.200	0.523
	Translation	20	0.000	0.000
SPW	No response	20	0.800	1.196
	Unclassifiable	20	0.300	0.470
	Phonological-formal	20	4.500	1.821
	Pseudo word	20	0.600	0.820
	Translation	20	0.000	0.000

Note: Each subject’s full score was 15 points

An One-way Repeated Measures ANOVA test was conducted to compare scores of different types of non-semantic responses in both the SPW and FPW. The means and standard deviations were presented in Table 5.12. With regard to FPW, there was a significant effect of response type (Wilks’ Lambda=.248, $F [4,16]=12.147$, $p<0.01$,

multivariate partial eta squared=.752). With regard to SPW, there was a significant effect of response type as well (Wilks' Lambda=.089, $F [(4,16)]=40.732$, $p<0.01$, multivariate partial eta squared=.911). Ranking of those responses in the SPW group from high to low was phonological-formal (72.58%) > no response (12.90%) > pseudo word (9.68%) > unclassifiable (4.84%) > translation (0%), and ranking of the FPW was phonological (76.36%) > unclassifiable (12.73%) > pseudo word (7.27%) > no response (3.64%) > translation (0%). In terms of percentage, phonological-formal responses were dominant, with 72.58 % of non-semantic connections in SPW, and 76.36% in FPW. This means that in the SPW-to-FPW process, the number of phonological-formal responses declined, but their percentage increased. The number of "no response" declined significantly, and its percentage declined as well.

5.4.4 Qualitative Data Analysis

Qualitative analysis was conducted as a supplement to quantitative analysis. All participants' responses were examined, so that some features could be detected. Through the examination it was found that in both the SPW and FPW groups, there were few responses that were obviously influenced by subjects' mother language, no matter in culture (e.g. spring → festival), or in expression (e.g. eat → vinegar). This may somewhat correspond with the phenomenon of zero translation response in both SPW and FPW group, which indicated that to upper-intermediate to advanced L2 learners or to words with high familiarity, the L2 mental lexicon is stored independently from the L1 mental lexicon.

The second feature detected was that for both SPW and FPW, most "pseudo word" responses were not fake words created without any reasons; rather, most pseudo words were created with semantic or phonological-formal relations with the cues. For

example, the pseudo word “timous” may most likely be created to mean “timorous”, and it was semantically related to the cue word “cowardly”; “wreckle” was made by some testees, which was phonologically or formally related to the cue word “wreck”; “hostly” was created because it was phonologically or formally similar to the cue “hastily”. This phenomenon may imply that the testees had grasped a certain amount of linguistic and semantic knowledge on the target language, and with the knowledge they endeavoured to add nodes in their lexical network when there were empty positions.

5.5 Discussion

The results of VKS and word association tests are presented in Section 5.4. In this following section, these results are given further analysis and discussion, aiming to answer the three research questions proposed in Section 5.2. The final goal of answering these questions is to find explanations for the threshold phenomenon from cognitive perspective. The cognitive factors that may motivate SPW-to-FPW development are discussed as well.

5.5.1 Differences of Lexical Representation between SPW and FPW

The results of this sub-study answered the three research sub-questions proposed in section 5.2. It was found that there were a number of differences in the testees’ mental representations between SPW and FPW. First, the semantic and non-semantic distribution was different. FPW were found to have significantly more semantic nodes than SPW did, and had significantly fewer non-semantic nodes than SPW did. In terms of the proportion of semantic and non-semantic links in lexical network,

semantic nodes accounted for the majority in FPW, and non-semantic nodes constituted only a small part. By contrast, in SPW, although semantic nodes outnumbered non-semantic nodes, the advantage of semantic connections was not as evident as that in FPW. This could be reflected by the fact that both semantic and non-semantic connections accounted for nearly 50% in the lexical network.

Secondly, the two types of words' semantic distributions were different. FPW's paradigmatic and syntagmatic nodes significantly outnumbered those of SPW, but there was no significant difference in encyclopedia nodes between these two groups. Ranking of these responses in FPW from high to low was paradigmatic > syntagmatic > encyclopedia; ranking in SPW from high to low was paradigmatic > encyclopedia > syntagmatic. In addition, the proportion of syntagmatic and encyclopedia nodes in semantic connections between the two types of words was different. In FPW, the syntagmatic connections accounted for 17.01% of semantic connections, and in SPW, they accounted for 9.36%; the encyclopedia connection in FPW accounted for 9.13%, and in SPW it accounted for 18.71%. Moreover, both paradigmatic and syntagmatic distributions between FPW and SPW were different. In terms of paradigmatic distribution, the number of synonym and antonym nodes in FPW was significantly higher than those in SPW. Ingredient nodes in FPW were significantly less than those in SPW. In terms of syntagmatic distribution, there were significantly more syntactic nodes in FPW than those in SPW.

Thirdly, the non-semantic distribution in SPW and FPW was different. FPW had significantly less phonological-formal and "no response" nodes than SPW did. With regards to the percentage, phonological-formal nodes' percentage in non-semantic connections tended to increase from the SPW group to the FPW group, although the

amount of it tended to decrease. By contrast, both the amount and percentage of “no response” tended to decrease in the SPW-to-FPW process.

From the comparison, it could be concluded that the lexical networks of SPW and FPW were organised differently in many aspects, and some features reflected in the comparison was worth paying attention to. First, the tendency of “semantisation” in the SPW-to-FPW process seems to be salient. In SPW-to-FPW process, semantic nodes tend to increase until they become dominant, and non-semantic nodes tend to decrease until they lose their dominance in the lexical network. This adjustment seems to be helpful for word production. More semantic nodes in a word web may mean that there are more semantic paths available, which could lead language users to the target word when the word is needed in speaking or writing. Moreover, there are less non-semantic nodes, which block the access to the target word in word-searching process. Therefore, the target word will be accessed and retrieved more easily with the adjustment.

Secondly, in the SPW-to-FPW process, different types of nodes tend to develop at different rates and even in different directions. This feature can be reflected in a number of aspects. For example, paradigmatic, syntagmatic and encyclopedia nodes tend to grow at different degrees. Paradigmatic nodes grow to different extents, showing salient advantage over other types of semantic nodes. Syntagmatic nodes grow to a great extent as well, both in number and in percentage of semantic links. Encyclopedia nodes do not increase in number, and their percentage in semantic links decline. This imbalanced change results in the change of inner semantic distribution, where syntagmatic nodes surpass encyclopedia nodes in both number and percentage. The semantic distribution changed from paradigmatic > encyclopedia > syntagmatic in SPW to paradigmatic > syntagmatic > encyclopedia in FPW. It is believed in this

sub-research that this change is helpful for word production. According to the definition by Nissen and Henriksen (2006), syntagmatic responses are those that have syntactic or compound/chunk relations with the prompt word, and meanwhile they are semantically related to the prompt. Encyclopedia nodes are concerned with the prompt word's world knowledge, including cultural, social knowledge and personal experiences. Therefore, in the SPW stage, when words cannot be produced but can be recognised and comprehended correctly, encyclopedia connections seem to be more important than syntagmatic connections. This is because the target word may not need much grammar or compound/chunk information for comprehension, but need encyclopedia information to have a deeper and more comprehensive understanding of the word. However, in the FPW stage, the situation seems to be reversed. Syntagmatic knowledge becomes significant, as sentence construction needs syntactic and sometimes compound/chunk knowledge. In speaking and writing people usually produce words in sequence, such as in phrases or sentences, rather than produce isolated words that are unrelated to each other (Meara & Fitzpatrick, 2000). Without enough syntagmatic nodes in the lexical network, production of the target word in contexts will be difficult. By contrast, encyclopedia knowledge may not facilitate word production to the same degree. Therefore, in the FPW stage syntagmatic nodes seem to play a more important role than encyclopedia nodes do; the growth of syntagmatic nodes helps motivate word production.

It was noticed that the phenomenon of “syntagmatic-paradigmatic shift” did not occur in this sub-study. A number of previous studies (e.g. Ervin, 1961; Fu et al., 2009; Namei, 2004; Söderman, 1993) detected that there is a “syntagmatic-paradigmatic shift” in both L1 and L2 mental lexicon development. These studies found that with the development of mental lexicon, paradigmatic nodes tend to keep increasing,

whereas syntagmatic nodes tend to keep decreasing. The notion that there is a paradigmatic-syntagmatic shift in L2 mental lexicon development seems to have been widely accepted now (Khazaeenezhad & Alibabaei, 2013; Namei, 2004; Piper & Leicester, 1980; Söderman, 1993; Wolter, 2001). However, the shift failed to occur in this sub-research, since both paradigmatic and syntagmatic nodes' amount and percentage grew in SPW-to-FPW process.

Reasons for the inconsistent results need to be investigated. Upon examining previous research more closely, it was found that almost all those studies adopted testees' age or language proficiency level as independent variables (e.g. Fu et al., 2009; Namei, 2004; Söderman, 1993). For instance, Namei (2004) conducted a cross-sectional research to compare word associations by children of different ages. His study found the phenomenon of syntagmatic-paradigmatic shift. However, it may be too hasty to conclude from the study that the shift is a result of the progress of word acquisition. Rather, the shift may have resulted from testees' enlarging vocabulary size or world knowledge when their age or language proficiency progresses. Fu et al. (2009) conducted a longitudinal study to investigate mental lexicon development. In her study, 40 words that testees were unfamiliar with were selected to be target words. Forty sophomores in English major participated in the study, and they were asked to do word association tests for three times during a three-month period. The study found a slight "syntagmatic-paradigmatic shift" tendency. The original goal of the research was to observe the mental lexicon change in individual words' acquisition development. However, since the study lasted for three months, the informants' L2 proficiency and world knowledge may also develop in this period. Additionally, the target words' acquisition level or familiarity level seemed to be only judged by the researcher's subjective assumption that in the three months those words developed.

No measuring instrument was adopted to ensure whether they truly developed or not, and to what degrees they had developed. A slight “syntagmatic-paradigmatic shift” was discovered in the study, but the slight degree may have resulted from the short duration of three months, since in three months informants’ L2 proficiency and world knowledge may only develop at a limited degree. Namei (2004), Zareva (2007), and Zareva and Brent (2012) also believed that the growth of age and language proficiency would make informants elicit more paradigmatic reactions, since age difference may cause difference in testees’ world knowledge, and L2 proficiency difference may cause difference in testees’ vocabulary size. Therefore, the conclusion could not confirm that syntagmatic-paradigmatic shift does exist in words’ acquisition development. That is to say, the notion that there is a paradigmatic-syntagmatic shift in words’ acquisition progress in fact has never been confirmed by previous research. However, conclusions of those researches sometimes seem to mislead people that there is a syntagmatic-paradigmatic shift in word acquisition development (e.g. Wolter, 2001). People tend to overlook the fact that those studies indicating the shift adopted age or language proficiency, not word acquisition level, as the independent variable. In fact, the syntagmatic-paradigmatic shift has nothing to do with the individual words’ mental lexicon development. The misunderstanding may have negative effects, because it may make ESL teachers think that in vocabulary teaching they should focus more on paradigmatic nodes, rather than on syntagmatic nodes. This teaching principle may not be effective.

In addition, it seems that even the method of adopting age or language proficiency level as the independent variable is problematic. One of the evidences reflecting the problem is that a number of relevant studies ended in illogical results. For example, S. Zhang (2004) investigated some advanced Chinese ESL learners, and found that these

learners produced a great deal of phonological responses. Söderman (1993) discovered that advanced ESL learners produced numerous phonological responses, but beginners produced a number of paradigmatic responses. Namei (2004)'s finding is similar in that even the most advanced L1 and L2 speakers have phonological and syntagmatic reactions, whereas even the very beginning L1 and L2 users have a great deal of paradigmatic reactions. Söderman (1993) and Namei (2004) therefore pointed out that no matter in which proficiency level, there are always words that are unknown, partially known, or well known; each word would experience a sequence of development from being unknown to well known, and the sequence is not affected by language proficiency. In other words, it is the word's acquisition level, rather than proficiency level, that would determine response patterns. In research adopting language proficiency as the independent variable, conclusions may be completely different if the selected words have different acquisition levels. Similarly, studies adopting age as the independent variable may have the same validity problem as well. No matter in which age level, there are always words that are unknown, partially known, or well known; each word would experience a sequence of development from being unknown to well known, and the sequence is not affected by age, if the factor of cognitive ability in age is excluded. In other words, it is the word's acquisition level, rather than age, that would determine response patterns. In research adopting age as the independent variable, conclusions may be completely different if the selected words have different acquisition levels.

In this sub-study, word acquisition levels were adopted as the independent variable, which was named as "familiarity" and measured by VKS test. The variables of age and language proficiency were controlled, since the testees were of similar proficiency and age. It is believed that the sub-study's validity is comparatively

satisfactory. The “syntagmatic-paradigmatic shift” was not found in this sub-study. Therefore, it seems reliable to conclude that at least in the SPW-to-FPW stage, both paradigmatic and syntagmatic nodes keep increasing, and no paradigmatic-syntagmatic shift would occur.

The feature of the nodes’ imbalanced development is reflected in the different types of paradigmatic responses as well. In the paradigmatic responses, synonym got prior growth, and it was the largest type of response in paradigmatic connections. Antonym experienced significant growth as well, although it only accounted for a small proportion in paradigmatic connections. All other types of paradigmatic responses did not grow significantly. According to the Spreading-activation Model by Collins and Loftus (1975), the more properties two nodes share, the more related and connected they are to each other. Since synonym contains similar or identical meaning as the stimulus word does, it may be the closest type of node to the stimulus. Therefore, the prior development of synonym may reflect a tendency of building “shortcuts” in network construction in the SPW-to-FPW process. This tendency is helpful for word production, since more synonym connections may assist target words to be accessed and retrieved more quickly and easily. The feature of imbalanced development can be evidenced by the prior growth of syntactic nodes, and the significant decrease in phonological-formal nodes and pseudo word nodes.

The third feature is that lexical connections tend to get stronger in the SPW-to-FPW process. This feature can be reflected in a number of aspects. For example, in the SPW-to-FPW process, paradigmatic and syntagmatic connections had significant growth, but encyclopedia nodes stagnated in development. As it has been pointed out, Encyclopedia response (e.g. Antarctica → ozone hole) is related to testees’ world

knowledge. Although this type of response is semantic, it seems not as closely related to stimulus words as paradigmatic and syntagmatic nodes do. The semantic distance between the stimulus word and its encyclopedia response is comparatively distant, and the connection between them is loose. Therefore, the phenomenon that paradigmatic and syntagmatic got prior development may indicate that the lexical network tends to get stronger in the SPW-to-FPW process. The change in non-semantic connections can also demonstrate this feature. In non-semantic connections, the proportion of phonological-formal response kept rising, although the amount of it was shrinking. Other types of response either declined in percentage, or did not change in percentage. Compared with other types of non-semantic response, phonological-formal response is more related to the cue word, since it is phonologically or formally similar to the cue. Other non-semantic response, such as non-classifiable, pseudo word and “no response”, have no common properties in any aspect with the cue. Therefore, the increasing proportion of phonological-formal nodes in non-semantic connections suggests that non-semantic network tend to get stronger and tighter, although its size tends to shrink. Additionally, the increase of semantic nodes and the decrease of non-semantic nodes demonstrate this feature as well.

However, it may be necessary to clarify that the decrease of some types of nodes does not necessarily mean they disappear from the word web. Since the word association tests only asked testees to give the first three reactions to the cue, only the central area of the network was observed in this sub-study. Therefore, the decrease of some types of nodes may not mean that those nodes have been “wiped out” from the word web, but rather they remain their presence but are not at the center of the word web any more. For example, when the target word is SPW with high familiarity, there are

numerous encyclopedia nodes around the target word. These nodes offer the language user adequate social, cultural or personal information on it, and they help the language user have an in-depth comprehension of the word. When the target word develops to be FPW, the encyclopedia nodes are still in the network, but great deals of them move to less important positions, and their original positions are taken by syntagmatic nodes. According to Laufer and Goldstein (2004) and Meara and Fitzpatrick (2000), the sequence of word development is to become receptive first, and on this basis to develop into productive ones. Therefore, in the productive stage, nodes that function at the receptive stage are still functioning, but nodes that are more important for word production are integrated into the network.

Similarly, in the early acquisition stage, the phonological-formal nodes may be dominant in its lexical network, as reported by Fu et al. (2009). Fu et al. (2009) found that setting up phonological-formal connections with other words might be the first step in word acquisition. With further development, more semantic nodes are integrated into the web, and they replace phonological-formal nodes in the central positions. However, it does not mean that the phonological-formal nodes disappear from the network, but rather they move to less important positions. In other words, they do not play a role as important as before in word activation and retrieval. Their functions are still there, which is to offer the language user the target words' spelling and pronunciation information.

5.5.2 The Lexical Construction Procedure in SPW-to-FPW Process

The results of this sub-study have shed some light on how the lexical network is constructed in SPW-to-FPW process. When a word can be recognised with high familiarity, but is not ready to be frequently produced, almost half of its nodes in its

lexical network are non-semantic. These non-semantic nodes become “barriers” when L2 users try to access and retrieve the word’s meanings from mental lexicon. The other half of nodes are semantic. Most of the semantic nodes at this stage are paradigmatic and encyclopedic. These paradigmatic and syntagmatic nodes can offer L2 users information on the target word from various perspectives, especially on synonyms, antonyms, social and cultural knowledge, grammatical knowledge, and collocational knowledge. Therefore, the word not only can be correctly recognised, but also can be well comprehended by the language user. At this stage, there are not as many syntagmatic nodes as paradigmatic and encyclopedic nodes; therefore, the lexical network does not have adequate paths for L2 users to retrieve the target words’ syntactic or compound/chunk information for production.

With further development, semantic nodes keep growing, and non-semantic nodes keep shrinking, until semantic nodes obtain salient advantage over non-semantic ones, and they become the majority in the lexical network. At this stage, not only does the size of semantic connections get enlarged, but also the inner structure of semantic connections changes as well. Paradigmatic and syntagmatic nodes get significant growth, but encyclopedia nodes do not. The result of this change is that the syntagmatic nodes surpass the encyclopedia nodes and become the second biggest type of response. In paradigmatic connections, synonyms get prior development and become the majority in paradigmatic connections. Therefore, when target words are needed in oral or written expression, a large number of paradigmatic nodes (especially synonyms) and adequate encyclopedia nodes would offer numerous channels for target words to be reached and retrieved. Meanwhile, a great deal of syntagmatic (especially syntactic) nodes are accessed and activated, making the target word ready to be produced in oral and written contexts.

In sum, this part of study detected three features in the SPW-to-FPW process. The first feature is that lexical network tends to get semantised; the second feature is that different types of nodes tend to change at different speeds and in different directions (increasing or decreasing); the third feature is that connections of lexical network tend to get stronger and tighter. According to the Spreading-activation Model, mental lexicon is organised as a network of nodes; the nodes are connected with each other at “different distance, strength, and spreading-activation levels” (Liu et al., 2012, p. 60). In other words, links in lexical network have different accessibility, such as strength and travelling time (Collins & Quillian, 1969, 1972). The results of this sub-study, to some extent, confirm the hypothesis of the Spreading-activation Model in that at least in the SPW-to-FPW stage, the lexical network tends to develop in the way that nodes become more easily accessed, activated, and retrieved. Therefore, some words’ stagnation in the SPW-to-FPW development may be attributed to some problems in lexical network construction. It may lie in the inadequate semantisation, i.e., the inadequate growth of semantic nodes, especially synonym, antonym, and syntactic nodes. Therefore, in the SPW-to-FPW development, semantic construction seems to be important, and great importance should be attached to synonym, antonym and syntactic nodes in particular.

It may be necessary to reiterate that the conclusion from this sub-study does not mean that some other types of nodes are not important. Rather, those nodes have their own functions in lexical development. For instance, encyclopedia nodes can help language users obtain in-depth comprehension for the target word, and phonological-formal nodes can assist language users to grasp spelling and pronunciation of the word. The conclusion of this sub-study just means that at the stage when the target word has become SPW with high familiarity, numerous encyclopedia and phonological-formal

nodes have already been built and integrated in the lexical network. Learners have grasped the words' form and pronunciation, and have had an in-depth comprehension of the word. In this stage these nodes have completed their mission and withdrawn to less important positions in the lexical network. Therefore, these types of nodes should still be built and strengthened in vocabulary teaching and learning, but at earlier stages. In the stage when SPW aim to develop into FPW, more focus should be given to synonym, syntactic and antonym nodes. It seems that at different word acquisition stages, there are different types of nodes that need to be focused on.

5.5.3 Pedagogical Implications

Based on the results of this sub-study, some suggestions are proposed on ESL vocabulary teaching and learning. Firstly, semantisation is necessary in the SPW-to-FPW development. Therefore, after the target words become SPW with high familiarity, semantic building should be given adequate attention in ESL teaching and learning; more semantic nodes should be integrated in mental lexicon. This goal may be achieved by semantic input and output. The semantic input activities may include intensive reading, extensive reading, listening, etc. The semantic output activities may include speaking and writing. That is to say, adequate quantity of reading, listening, speaking, and writing should be ensured. Adequate activities in these areas may be the first step to motivate SPW-to-FPW development.

Secondly, semantisation alone may not be enough. The results of this sub-study indicate that some types of nodes seem to play more important roles than the others in the SPW-to-FPW process. These nodes include synonym and syntactic nodes. Antonym nodes play a somewhat positive role as well. In other words, not only the

amount of input and output is significant, what to input and output is important as well.

Therefore, after SPW have reached high familiarity, more attention should be focused on the building of synonym, syntactic and antonym nodes. This sub-research found that there are several activities that may be helpful in building synonym nodes. For instance, paraphrase may be an efficient way, since paraphrasing involves target words being substituted by words of similar meanings. Other activities include rewriting and retelling, which ask students to have oral or written restatement. Brainstorm can be used as well (it can also be used to collect antonyms), where students are asked to collect synonyms of a target word as many as possible.

With regard to syntactic nodes, several tasks may be helpful. Firstly, after words become SPW with high familiarity, adequate grammar input should be offered by teachers. For example, if the target word is a transitional verb, the objects that can be followed, the subjects that can be used before it, and the adverbials that can co-exist with the verb should be provided as well. Sentences which can demonstrate grammatical features of the target word could be given as examples. Moreover, various types of grammatical practices are encouraged to be carried out, such as matching (which asks students to connect the target word with its correct usage), True or False (which asks students to judge whether the word is used correctly in a sentence), and filling in the blanks (which asks students to use the target word with correct syntactic forms). In addition to input of grammar, output of grammar should be encouraged. This part of study recommended that students be asked to make sentences with the target words. If syntactic mistakes occur in the sentence, feedback should be provided in time, and mistakes should be corrected. With these measures, syntactic nodes may obtain more chances to be built in lexical network, and those

nodes that already exist may get more chances to get strengthened and become closer to the target word.

Students' consciousness seems to be necessary as well. They should be advised to be more conscious on semantic construction. They should be suggested to pay enough attention to the target words when they encounter those words in contexts, and to consciously use them in speaking and writing. In addition to the methods suggested in this sub-study, there may be other teaching and learning methods that can facilitate lexical construction. Teachers and learners are advised to design different activities based on the principles detected from this sub-study to promote FAV development.

5.6 Conclusion

This chapter compared lexical network representation between SPW and FPW. The results indicate that the two types of words' representations differ in numerous aspects, including semantic and non-semantic distribution, semantic distribution, and non-semantic distribution. Three features are revealed from the results. Firstly, the lexical network tends to be semantised in SPW-to-FPW development. Secondly, nodes of different types develop at different extents and directions; some types of nodes seem to play more important roles than others in SPW-to-FPW process. Thirdly, connections in lexical network tend to get stronger and tighter in the process. The results of this sub-study to some degree confirm the hypothesis of the Spreading-activation Model that nodes tend to be more easily accessed, activated, and retrieved in lexical network development. Based on the conclusion, it is proposed that one of the chief reasons for the threshold phenomenon is that some SPW with high familiarity do not get adequate semantisation in the SPW-to-FPW development. In

particular, synonym, antonym and syntactic nodes' inadequate growth may play a critical role for the stagnation.

In this chapter the word association sub-study is presented. The purpose of the sub-study is to explore the mental lexicon construction procedure in SPW-to-FPW process, and to detect effective ways to promote FAV development. In the next chapter, the qualitative sub-study of interview is presented. The purpose of the sub-study is to find possible factors in learning and teaching that may affect FAV development, and to find possible ways to facilitate FAV development.

Chapter 6

Interview

6.1 Introduction

This chapter reports on the investigation of students' consciousness on free active vocabulary (FAV) development, teachers' guidance on free active vocabulary learning, the effective learning strategies students adopt (if any) to enlarge FAV storage, and the effective teaching strategies (if any) teachers adopt to assist students' FAV development. The aim of the sub-study is to find possible factors in learning and teaching that may affect FAV development, and to find possible ways to facilitate FAV development. The chapter first poses the aspects around which interview questions were designed, then presents research method, including instruments, data collection, data preparation and data analysis. At the end of the chapter, results are reported and discussed, and conclusions are made.

6.2 Aspects of Interview Questions

The interview questions were designed surrounding four aspects:

- (1) Students' awareness of FAV in L2 study;
- (2) Learning strategies (if any) adopted to improve FAV;

- (3) Teachers' guidance on FAV learning; and
- (4) Teaching strategies (if any) that may help promote students' FAV development.

6.3 Research Method

This section presents major components of the research method adopted in the sub-study, including instruments, procedure of data collection, data preparation, and data analysis. The participants of the interview have been introduced in Chapter 3.

6.3.1 Instrument

The instrument adopted in the sub-study is a face-to-face interview. Questions in the interview are designed in four aspects, which are presented in Table 6.1. It should be noted that in the formal interview, extra questions may be added when necessary, so that more detailed information on the relevant topic can be obtained.

Table 6.1. *Aspects and Corresponding Questions in the Interview*

Aspects	Corresponding questions
Aspect One: Students' consciousness of FAV in L2 study	<ol style="list-style-type: none"> 1. Have you ever had difficulties in searching for suitable words to express the meaning you want to express in L2 writing or speaking? 2. What are the difficulties (if any)? 3. Do you know free active vocabulary? Or do you know this type of vocabulary knowledge, even if you do not know the terminology of it? 4. Have you ever attempted to improve your free active vocabulary size in L2 study?
Aspect Two: Learning strategies adopted to enlarge FAV	<ol style="list-style-type: none"> 1. Have you ever taken some measures to enlarge FAV size? 2. What are the measures (if any)? 3. Do you think these measures are effective?
Aspect Three: Teachers' guidance on FAV in L2 teaching	<ol style="list-style-type: none"> 1. Have your teachers told you that FAV is important?

	<ol style="list-style-type: none"> 2. Are there many teachers who did so? 3. Have any teachers given you any suggestions on how to enlarge FAV? 4. What suggestions did they give you (if any)?
Aspect Four: Teaching strategies that may help students' FAV development	<ol style="list-style-type: none"> 1. Do you think some teaching strategies are helpful in FAV acquisition? 2. What are they (if any)?

6.3.2 Data Collection Procedure

6.3.2.1 Pilot Study

A pilot study was conducted before the formal interview. Five year-three students who were studying in the same major and from the same university took part in it. Interview questions were presented, and feedback was collected. The pilot study aimed to ensure that the interview questions were clear enough to interviewees, and to check whether it was necessary to design additional questions in the formal interview. It also aimed to find the suitable length of time to conduct the interview. The interviewees' answers and feedback showed that the designed questions were clear to interviewees. However, it was found that some students did not know the terminology of free active vocabulary, but they in fact knew that there was this type of vocabulary knowledge. Therefore to avoid drawing wrong conclusions, an explanation on what is free active vocabulary was to be provided to the interviewees, and a follow-up question was added to the original question of "do you know free active vocabulary". The added question was "do you know that there is this type of vocabulary knowledge, even if you don't know the term 'free active vocabulary'?" The pilot study showed that the time needed for each interview was around 5 to 7 minutes.

6.3.2.2 Formal Data Collection

The formal data were collected in June, 2016. It was considered to be a suitable time to collect the data, as at the end of the four-year university study, students would have enough spare time to participate in the interview. In addition, it was a time when they were able to have a comprehensive reflection of the learning and teaching activities in the past four years. Before the data collection, the student investigator asked the interviewees to answer interview questions as detailed as possible. They were also encouraged to raise questions if they had any doubts on the interview questions. Then the interview started. Each student was interviewed separately, so that their answers would not be affected by other interviewees. The entire process was recorded.

6.3.3 Data Preparation

Interviews were audio recorded. Then the recordings were played and checked by the student investigator to ensure that the interviewees answered the questions seriously. If details were given in the interviewees' answers, it was believed that the interviewees answered the questions seriously. It turned out that all the interviewees did the interview with a serious attitude, and therefore all the data were valid. Then the recordings were transcribed by the student investigator for further analysis. In order to protect privacy, the interviewees were kept anonymous, and they were coded as Interviewee A, Interviewee B, and Interviewee C.

6.4 Results and Discussion

Transcription of the recording consists of approximately 3,260 words in total. The student investigator read the transcription sentence by sentence, and then had the

transcription encoded, arranged, and categorised. The interview results are reported and discussed in this section.

When being asked “have you ever felt difficulties in searching for suitable words to express the meaning you want to express in writing or speaking?”, the interviewees’ answers were almost the same. They reported having difficulties in seeking for the right words to use in writing and speaking. Specifically, the students’ challenge primarily lied in the fact that sometimes their effort in searching for the most accurate word failed. Interviewee A felt unsure sometimes about the precision of the candidate words in writing, or sometimes he could not find an exact word to express what he wanted to express. In this condition, he had to use a substitute instead. This led to the loss of accuracy in his writing. Interviewee B had the same concern. She said:

I definitely feel the difficulty. As a senior student, I start to have job interviews. In some of the interviews, the interviewers would require me to have oral communication (in English) with them. Sometimes I feel my vocabulary is rather limited... The word that I use is not precise, but I cannot find the precise one. (Interview B)

Interviewee C told the investigator that she suffered from this trouble frequently:

I think words have both general meaning and specific meaning, but in many cases I only know the general meaning. Thus, in a specific circumstance, I may not be able to find the specific word, and thereby have to use a general word to replace it.

The student then gave an example. Once she wanted to express “the ‘American dream’ contains different potential meanings in different times”, but she could not find a suitable word to express “potential meaning”, such as “connotation” or “implication”. Thus she had to use the word “meaning” instead. She said this problem

often occurred. Moreover, she told the investigator that the difficulty she faced was not in choosing the right word among options. The trouble seemed to be even worse: she could not think of optional words at all. She had to use a basic, general word, as it was the only one that could come into her mind. This difficulty corresponds with the statement by Swain (2000, p. 100): “Learners may notice that they do not know how to express precisely the meaning they wish to convey at the very moment of attempting to produce it...they notice, so to speak, a ‘hole’ in their interlanguage.” This problem may result in the loss of important information the student wants to convey in writing and speaking.

It can be inferred from the interview that the difficulty in FAV usage is common even among upper-intermediate to advanced students. These students’ difficulty is primarily in retrieval, rather than in the selection of the right word among retrieved words. Findings in Chapter 5 may shed some light on the retrieval difficulty. As is concluded in Chapter 5, the lexical network is organised differently between frequently produced words (FPW) and seldom produced words (SPW). The semantic nodes of FPW tend to outnumber those of SPW, whereas non-semantic nodes of SPW tend to outnumber those of FPW. With regard to proportion, semantic links account for the majority in FPW, but in SPW they do not. Taking a closer look, it can be seen that FPW’s paradigmatic and syntagmatic links are significantly more than SPW’s. Within paradigmatic links, FPW have significantly more synonym and antonym nodes than SPW do. Within syntagmatic links, FPW have significantly more syntactic nodes than SPW do. The investigation in Chapter 5 also found that there is difference between FPW and SPW in the number of non-semantic links. There are significantly less phonological-formal and no-response nodes in FPW than in SPW. The finding seems to indicate that SPW may not have enough semantic nodes, but too many non-

semantic nodes in their lexical network. Specifically, SPW may not have adequate synonym, antonym, and syntactic nodes, but have too many phonological-formal and non-response nodes around them. The lexical representation sheds some light on reasons for the retrieval difficulty. Those SPW may not have enough semantic paths to be reached in lexical network, but have too many non-semantic paths to block the retrieval. In addition, the lack of syntactic links makes it hard to be produced in context, as grammatical knowledge is an inevitable factor in sentence creation.

The next set of questions is on students' consciousness of FAV in L2 study. All the interviewees indicated their awareness of FAV. However, when they were asked whether they had made effort to enlarge FAV storage in L2 study, two of them answered that they only took "some" effort to it. For example, interviewee B reflected that for a period of time she tried to enlarge FAV, but gave up later. This indicates that some students know the existence of FAV, and have attempted to improve it, but the attempt seems to be inadequate, as they may give up after a while.

Effective learning strategies in FAV development were then investigated. Two interviewees reflected that taking deliberate "output" practice was helpful to them. For example, Interviewee A reflected that he tried to pay extra attention to the target word, and then forced himself to produce the word. When he encountered a new word in study, he would deliberately create chances to use the word in writing or speaking. Interviewee B stated a similar strategy. In reading, if some new words attracted her attention, the student would put them down in a notebook. If these words were encountered for a second time, she would pay more attention to it, and attempted to apply these words in writing. She said the purpose of this activity was to help these words be transferred from receptive to productive. The student stressed that if the word was not used, it would be forgotten with time passing by.

Analysing the context and putting down example sentences were deemed to play positive roles as well. Interviewee C told the investigator that in reading, when she first encountered a new word that drew her notice, she would put down the word. If she encountered the word again, she would pay more attention to it, and considered why it was used in this context. By this method she would learn how to use the word. If she noticed that a word appeared frequently, she would look it up in a dictionary, and then copied the example sentences to her notebook. The reason to do this was because these examples might be helpful for her writing.

In the interview one phenomenon attracted the investigator's attention: the students seemed to be sensitive to word frequency in FAV acquisition. Two interviewees stated that usually if they only encountered the word once, they might notice it, but would not consider how to use it. Only when they met the word repeatedly would they try to output it by themselves. This tendency may imply that word frequency is considered to be important by students. If a word appears frequently, they will think that the word is worthwhile to be transferred into FAV; if a word does not appear frequently, they tend not to think so, and thereby will not make efforts to develop it into FAV. This finding corresponds with the finding in the 12-month longitudinal sub-study that words are usually acquired in order. More frequently used words are usually learned before less frequently used ones (Schmitt et al., 2001). The suggestion is that repeated occurrence of the target words may be helpful for FAV development. In texts or lectures, teachers are advised to deliberately increase the occurrences of the target word, so that students' attention can be attracted. This measure may have an indirect positive effect on FAV development.

The interviewees described some other strategies that they felt effective as well. One strategy was collecting synonyms. Interviewee C introduced this method, and she

thought it was very helpful to her. She said at lower proficiency level, words encountered were mostly basic words. With further study, more synonyms were encountered in intensive and extensive reading. The interviewee deemed that collecting synonyms was helpful to writing. In this way various words could be used, therefore the writing would not be as plain as water. The student pointed out that if only the most basic words were applied, quality of the writing would be similar to that of middle school students.

Another strategy being reported was to make use of “context”. Interviewee A told the student investigator that he often looked for authentic reading materials, such as those in foreign English magazines. If he encountered a new word in these reading materials, he would try to figure out the meaning of the word in the context. Then he would know that it was appropriate to use the word in this type of context. When he met a similar circumstance, he would know that it was proper to use this word in it. The interviewee reported that this learning strategy worked for him.

The Output Hypothesis by Swain (1985, 1995) can be used to explain why this strategy is effective. The Output Hypothesis states that although comprehensible input is essential in second language acquisition, solely having input is not enough. Comprehensible output is essential as well. The hypothesis holds that language production, no matter in spoken form or written form, would elicit language learning/acquisition to occur (Swain, 1985). According to Swain (1985), being “pushed” not only means to make learners convey a message, but also to convey it precisely, coherently, and appropriately. He believes that the act of “pushing” would force ESL learners to make more mental effort in study, thus process the language more deeply. Swain’s (1985) Output Hypothesis is originally on general second language acquisition, rather than specifically on vocabulary or FAV acquisition.

However, some scholars found this theory was relevant to L2 FAV learning as well, and therefore they applied this theory to L2 FAV study. For instance, Laufer (1998) holds that if L2 learners are not pushed to use L2 words, the words may never be activated, and they may only remain in passive vocabulary. Henriksen (1999) believes as well that pushed output helps promote receptive words to develop into productive words. Empirical studies have been conducted. Ellis and He (1999) found learners with output practice improved significantly in both receptive and productive vocabulary (FAV in this case) learning. The experiment by Fuente and José (2002) shows that negotiation of input plus pushed output resulted in more L2 target words' production than negotiation of input without pushed output. The experiment by Arash (2016) reveals that when L2 learners are given chances to produce the target words, their performance of word production is better than the performance when they are not given the chance. Arash (2016) therefore concluded that pushed output can improve L2 word production.

It may be necessary to note that the pushed output in this study refers to the pushed "semantic" output, rather than pushed "formal" output, as output of forms does not do much for the words' semantic construction. Barcroft (2006) administered an experiment to check the effect of forced output of forms on vocabulary learning. In his study, subjects attempted to learn a group of L2 words by viewing word-picture pairs. In the process, the participants were asked to copy half of the target words, but not to do so for the other half of the target words. The productive vocabulary knowledge was tested after the treatment. Subjects were required to write down words after viewing relevant pictures. It is deemed in this study that the activity of picture-word writing to some extent assesses subjects' FAV, since the test does not provide any hint on the words' spelling or pronunciation to subjects. The subjects had to

“freely and actively” produce the word when they saw the picture. Results of this study revealed that the no-copying group significantly outperformed the copying group in FAV learning. The implication of the study, according to Barcroft (2006), is that the forced output without access to meaning would have negative effects on vocabulary (FAV in this case) learning, as learners have to spare processing resource to encode forms of the words.

Although Krashen (1998), who proposed the Input Hypothesis, doubts the significance of output in second language acquisition (SLA), this sub-study seems to support the opposite view that output does play a positive role in SLA, specifically in FAV acquisition in this case. When students try to create sentences with the target word, they are “pushing” themselves beyond their current level of word use, as they force themselves to use a word that they never used before. They tend to make more mental effort in this condition than when they solely receive input of the word from the teacher. This may be the reason why the pushed output activity can work effectively for FAV development.

All the effective learning strategies reported by the interviewees can be explained by theories of Depth of Processing, Task-induced Involvement, and findings in mental lexicon. The Theory of Depth of Processing is proposed by Craik and Lockhart (1972). According to the theory, the depth at which a piece of information is processed is an important factor in memorization: the deeper the processing is, the better the memorization will be (Craik & Lockhart, 1972). For example, trying to memorize a word out of context and trying to memorize a word in context may involve different depth of processing: memorizing the word in context involves more meaning analysis than memorizing it out of context, and therefore it may bring better results (Yue, 2008). The shallow depth of processing might primarily be on sensory

levels, such as visual or acoustic processing; deeper processing might primarily be on meaning analysis and structure analysis (Craik & Lockhart, 1972). The strategies of making sentences, referring to a dictionary, taking notes on word meanings, putting down example sentences, analysing the context, and collecting synonyms may all result in a deep processing on the target word. Therefore, these learning methods work effectively for FAV development.

Another theory is Task-induced Involvement Theory, proposed by Laufer and Hulstijn (2001). Laufer and Hulstijn (2001) hold that different tasks tend to elicit different amounts of cognitive involvement load. The cognitive involvement includes three aspects: need, search, and evaluation. Different tasks contain different loads of involvement in the three aspects, thus result in various acquisition results. For instance, in the study by Duan and Yan (2004), experimented words were presented to participants in different ways. The participants of Group 1 read an article with target words and then did multiple-choice tasks, in which correct meanings for the target words should be chosen. Each of the tasks had more than one correct answer. The participants of Group 2 read the same article with the target words, and then did multiple-choice tasks for correct word meanings as well, but each task had only one correct answer. Group 3 read the article with the target words, and the word meanings were given to them directly. The three tasks contained different levels of involvement. The first task elicited all the three involvements. There was “need” to seek the word meanings, “search” to look for the correct choices of word meanings, and “evaluation” to judge whether the choices were correct or not. The second task elicited all the three involvements as well, but the amount of “search” and “evaluation” was most likely smaller than that in the first task. It was because in the second task only one correct choice needed to be selected and evaluated. By contrast,

in the first task two or more than two correct choices needed to be found, therefore it needed more effort in “search” and “evaluation”. The study showed that the first task resulted in the best word acquisition outcome, the second task came next, and the third one resulted in the least acquisition. The study verified the Task-induced Involvement Theory.

In fact, the Theory of Depth of Processing and Task-induced Involvement are related to each other, as the Task-induced Involvement is developed on the basis of Theory of Depth of Processing (Yue, 2008). The weakness of the Theory of Depth of Processing is that the “depth” cannot be quantified and measured. By contrast, the Task-induced Involvement overcomes this weakness, as the theory makes the amount of involvement possible to be measured and evaluated by investigating whether the task contains “need”, “search”, and “evaluation”. Therefore, the Task-induced Involvement makes relevant empirical studies feasible to be conducted (Yue, 2008).

Moreover, it is found that the effectiveness of the learning strategies can find explanations in mental lexicon construction. As is reflected in previous paragraphs, frequently produced words have a rich connection of semantic nodes. The strategies of analysing context, looking for word meaning in dictionaries, putting down example sentences, and collecting synonyms all help set up semantic nodes in the lexical network. The construction plays a positive role in transferring the target words into FAV.

The next series of questions aim to investigate teachers’ guidance on FAV learning and the effective teaching strategies they adopted to assist students’ FAV development. When being asked whether their teachers had told them the importance of FAV, all the interviewees answered yes. However, they recalled that only one teacher teaching them Intensive Reading in the third year did so. This indicates that

many teachers may not be conscious of the significance of FAV; even if they have the consciousness, they may not take any action to help students improve FAV. The guidance that students can obtain from teachers, therefore, is limited. In addition, teachers may lack knowledge on how to help students effectively, as the literature review reflected that the relevant research is rather limited. This condition makes students unable to get adequate guidance from teachers.

The lack of strategy guidance can be evidenced from the interview. When the investigator asked the interviewees whether their teachers had given them suggestions on how to improve FAV, two of them reported no specific methods recommended by their teachers. Only one interviewee reflected that the Intensive Reading teacher suggested them to put down new words in notebooks, and to deliberately use them in writing. The teacher told them that this was a method used by one of her former students, who had good writing capability.

It was found that few teaching strategies had been adopted to enlarge students' FAV. However, the interviewees described some teaching strategies that aimed for other purposes but played a positive role in FAV acquisition. For instance, two interviewees recalled that some teachers often provided them with synonyms of the target words. For instance, one of them reflected that once in class when students were studying the words "strength" and "weakness", the teacher provided them synonyms of "advantage" and "disadvantage", and "merit" and "demerit". In this way, not only the target words were learned, their synonyms were learned as well.

One of the teachers took a step further. Instead of solely providing synonyms, the teacher conducted "negotiation" to students. According to Arash (2016), "negotiation" means to negotiate the meaning of the word, such as to clarify, confirm, or repeat the word meaning to students. The study by Arash (2016) shows that input

of word meaning with negotiation leads to greater FAV gains than input of word meaning without negotiation. Another study testifies the positive effect of both negotiation and pushed output. The research by Teng (2015) shows that extensive reading with negotiation and pushed output activities prompt more FAV gains than solely extensive reading does. In this sub-study, the interviewee said:

Maybe because the teacher majors in literature, she has a precise understanding of words. Most teachers would only provide us with a general explanation on a target word, or just give us a translation or an example sentence with it. But she would analyse why this word is used properly here, and why the word is a good choice. She often does so.

This method goes one step further than offering synonyms, translations, or examples, as it gives students detailed information on the words' connotations, and a deeper analysis on the words' context. In other words, it provides students with "negotiations" of the word. Negotiation helps students have a deeper processing on the word. Negotiation may help students set up semantic network in mental lexicon as well. When students encounter a proper circumstance, the word will be more easily reached and retrieved from their mental lexicon. This may be the reason why negotiation can help students in FAV development.

6.5 Conclusion

The interviews in this chapter obtain information in the four aspects proposed in Section 6.2. The first aspect is students' consciousness to FAV in L2 study. It shows that students know FAV, and the awareness is most likely aroused by the difficulty they encounter in word production. Being aware of FAV's significance, they have taken some effort trying to enlarge it in study. However, students' effort may not be

adequate, as the number of strategies they use is rather limited, and they may give up the effort later. In summary, students may know the concept of FAV, but their effort to improve FAV may not be enough.

The second aspect is on effective learning strategies students adopt for FAV development. It is reflected that students do take some measures trying to enlarge their FAV size. The strategies that they feel effective include making sentences with the word, taking notes of the word's meaning, putting down example sentences, attempting to use the word in proper context, and collecting synonyms. A problem was revealed from the interview. The number of strategies adopted is limited, as each interviewee only reported one or two strategies, and confirmed later that those strategies were all that they used. The inadequacy in strategy use may be one factor for FAV stagnation.

The third aspect is teachers' guidance of FAV learning. Regrettably, the interview indicates that many teachers may not be aware of the significance of FAV; even if some of them know it, they fail to give adequate guidance to students on FAV learning. In the interview it was reflected that almost no teachers, except for one Intensive Reading teacher, had ever mentioned FAV to students. In addition, few suggestions had been given to the interviewees on how to enlarge FAV storage.

The fourth aspect is effective teaching strategies that may help students in FAV development. The interview shows that the guidance provided by teachers on FAV development is not adequate. However, some teaching methods were found by interviewees to be helpful to FAV development, although they were not adopted for this purpose. These methods include providing synonyms, negotiating word meanings, analysing contexts, arousing students' attention to target words, etc. Some effective teaching strategies overlap with students' learning strategies. Like the

effectiveness of learning strategies, the effectiveness of these teaching strategies can be theoretically explained by the Depth of Processing Theory, Task-induced Involvement Theory, and findings from the mental lexicon sub-study.

Some problems are detected from the interview. The first problem is that students lack proper guidance in FAV development. The interview shows that many teachers almost never mention FAV in teaching, let alone provide students with instructions and suggestions on how to improve it. Moreover, few teaching strategies are adopted to help promote students' FAV development. However, it may be hasty to blame this neglect on teachers, as research in this field is rather limited, and it cannot give teachers enough guidance yet. Although researchers have detected the FAV threshold phenomenon in L2 study, there is a limited number of studies, especially empirical ones, on how to tackle this problem. Even if teachers feel the great necessity to help students enlarge their FAV, they can find little reference to give them valuable suggestions. Therefore, relevant research is in dire need. On the students' side, realizing the significance of FAV, they have to "grope" for effective learning methods by themselves. The effectiveness of their trials thus mostly depends on luck. Some strategies may turn out to be helpful, some may not; some may be helpful, but may only be helpful to a specific person, rather than having a universal value.

The second problem is that although the interview shows that students know FAV, their effort to improve FAV seems not adequate. The inadequacy can be reflected by two aspects. Firstly, the number of strategies they use is rather limited. Each interviewee only reported one or two strategies that they felt effective. Secondly, some students may make effort to enlarge FAV, but tend to give up later. This may be because without teachers' guidance, students tend to get discouraged or forget to keep doing it.

Findings of the interview shed some light on ways to enlarge FAV. On the learning side, students not only need to know the significance of FAV, but also need to know useful strategies to tackle the problem. The interview shows that pushed-output activities, such as making sentences with the target word, may help. In addition, collecting synonyms, analysing context, taking notes of example sentences, paying extra attention to the target word may help as well. Continuous effort should also be made, rather than non-sustained efforts, as achieving significant FAV growth may need a long process. On the teaching side, teachers should realize the significance of FAV in L2 study. They should remind students to keep paying attention to FAV development. They should give students suggestions and guidance on how to tackle the stagnation problem. Moreover, teachers should adopt effective teaching strategies to assist students. The interview shows that helpful teaching strategies include negotiation, context analysis, synonym collection, and drawing students' attention to target words. It can be seen that some of the effective teaching strategies overlap with the effective learning strategies.

However, it should be noted that the effectiveness of those strategies is based on the interviewees' subjective judgment, not evidenced by empirical studies yet. Effective learning and teaching strategies not only should be guided by experience and theories, but also need to be tested by empirical studies. Therefore, it is recommended that more empirical research be conducted to verify the effectiveness of these strategies. It is believed that with quantitative support the conclusion will be more convincing and instructive to FAV learning.

Chapter 7

Documentary Analysis

7.1 Introduction

As English is a foreign language in China, formal class instruction is a major source from which students obtain English training. Therefore, the course structure plays a significant part in L2 learning. This chapter investigates the discipline-related courses students took in the period of the 12-month longitudinal sub-study. The aim of this sub-study is to investigate whether there are factors in course structure that may affect students' free active vocabulary (FAV) development. The chapter first poses research questions, then presents research method, including data collection, data preparation and data analysis. At the end of the chapter, results are reported and discussed, and conclusions are made.

7.2 Research Questions

This part of research continues to explore the following question:

- What are the possible factors in teaching that may affect FAV development?

7.3 Research Method

7.3.1 Data Collection

As the 12-month longitudinal sub-study started in May, 2015 and ended in May, 2016, the participants of the sub-study went through three semesters during the period. Therefore, the course structure in the three semesters was collected. The course structure included all the units students took in the three semesters and the number of hours per week for each unit.

7.3.2 Data Preparation

After the course structure data were collected, the non-discipline-related units, such as Politics, were removed, as they were not related to English language study. The discipline-related units that were not related to English language study, such as Japanese as a Second Foreign Language, were excluded as well. After the treatment, only those units related to English language were left for final analysis.

7.4 Results and Discussion

The units for final analysis are presented in Table 7.1.

Table 7.1. *Discipline-related Units*

Semester	Time	Units	Hours per week
Semester 1	May, 2015-July, 2015	Intensive Reading	4
		Introduction to Linguistics	2
		British Literature and Anthology	2
		English Pedagogy	2
		Theory and Practice of Translation	2
		Phonemics	2
		Research Methods in English	2

		Teaching	
		Selected Reading in English Essays	2
Semester 2	September, 2015-January, 2016	Introduction to Linguistics	2
		American Literature and Anthology	2
		Theory and Practice of Translation	2
		Selected British and American Newspaper Reading	2
		English Pedagogy	2
		English Academic Thesis Writing	2
Semester 3	February, 2016-May, 2016	Undergraduate Thesis	2

It is observed that in Semester 1 and Semester 2, the courses were concentrated on English training. According to the Syllabus of English Courses for College English majors by Chinese Ministry of Education (2000) in China, the discipline-related units for English major fall into three types. The first type is of professional skills, such as units for writing, reading, listening, etc. The second type is of professional knowledge, such as units on linguistics, literature, etc. The third type is professional-oriented, such as units on English teaching, business negotiation in English, etc. It is observed that the courses in the three semesters cover all the three types of units. For example, Intensive Reading and Theory and Practice of Translation are units of professional skills; Introduction to Linguistics, British Literature and Anthology, and Phonemics belong to units of professional knowledge; English Pedagogy and Research Methods in English Teaching are professional-oriented units.

It is also observed that there are only 3 units for output training. These units include English Academic Thesis Writing, Undergraduate Thesis Writing, and Theory and Practice of Translation. According to the requirement of The Syllabus of English Courses for College English Majors, one of the major aims of Theory and Practice of Translation is that through the unit students will be able to translate different types of essays between L1 and L2. In the unit, a great deal of translation practice is conducted, and the translated work is discussed and evaluated by the teacher and students. The Undergraduate Thesis Writing is another unit for output training. The entire unit takes about six months, involving several steps. Firstly, students will be asked to choose an academic topic that they wish to write about. The topic should be related to linguistics and applied linguistics, literature, culture, or translation. They should choose a teacher to be their supervisor according to the topic. Then students will meet their supervisor and discuss the appropriateness of the chosen topic. Usually when the topics are decided, a department meeting will be held, where the appropriateness of the topics are further examined and discussed. After that the research proposal will be made. Then the outline of the thesis will be written by students. The draft of the outline will be revised by the supervisor for at least three times, until it becomes satisfactory. Next, thesis writing of about 5,000 words is conducted. The draft of the writing is again revised by the supervisor for at least three times, until it becomes satisfactory. When the thesis is completed, the defence will be administered in school. If students pass the thesis defence, the final version of the paper, together with all the drafts of outlines and thesis writings, will be submitted to school for storage. Another unit for output training is English Academic Thesis Writing. It is a unit teaching students how to write an academic thesis in English. Usually in the unit the writing practice will be conducted, and the work will be

checked and evaluated by the teacher. These units are considered to be output-focused, as the teaching and training of writing constitutes a major part in all the three units.

It seems that the number of units for output training is in a salient disadvantage in the course structure. There are in total 15 discipline-related units, but among them there are only 3 units for output training. Meanwhile, the participants of the longitudinal sub-study received a large amount of L2 input from classroom instruction. As the participants are senior students in their last year of university study, the input covers a variety of professional topics. This feature can be reflected from The Syllabus of English Courses for College English Majors. For example, according to the syllabus, the aim of Selected British and American Newspaper Reading is that students will be able to read British and American newspapers and magazines. The syllabus requires that the teaching material of the unit should cover a wide variety of topics, and reach a certain level of difficulty. The teaching material should include comments on current events, editorials, political comments, special news reports, etc., from British and American mainstream newspapers and magazines. The topics should cover society, politics, economy, wars, environment, people, international relations, science and technology, etc. For another example, the syllabus stated that the aim of British Literature and Anthology is to foster students' capability to read, appreciate and understand the original English works, and to grasp basic knowledge and skills to have literature critiques. With regard to Introduction to Linguistics, the syllabus points out that its purpose is to let students know human beings' great achievements in language research, and to enhance students' understandings of languages' significant value to society, humanity, economy, science and technology, and personal cultivation. The syllabus requires that the unit should cover topics such as

language and mind, structure of language, origin of language, language and brain, language transition, etc.

Being a regular public university, the university in this study follows the syllabus in course design and teaching material selection. As the syllabus requires that the discipline-related units for seniors should reach upper-intermediate to advanced L2 proficiency level, it is rational to postulate that the input students receive from these units contains a large amount of words beyond the first 1,000 most frequent words. In other words, students have numerous chances to encounter vocabulary that are beyond the first 1,000 most frequent words from the input of these units. However, the longitudinal sub-study in Chapter 4 shows that no significant FAV growth occurred during the 12-month period of time.

Krashen (1982) proposed Input Hypothesis in second language acquisition (SLA). The Input Hypothesis holds that we acquire language “only when we understand language that contains structure that is a little beyond where we are now” (Krashen, 1982, p. 21). At the beginning the hypothesis was on general SLA. In 1989, Krashen extended it to L2 vocabulary acquisition, claiming that L2 input, such as reading and listening, would facilitate incidental vocabulary acquisition. He referenced a considerable number of empirical studies to support this view. From these studies of incidental vocabulary acquisition, Krashen concluded that “comprehensible input alone can do the entire job for vocabulary and nearly the entire job for spelling” (Krashen, 1989, p. 448). It is noticed that Krashen (1989) only mentioned the effect of comprehensive input on passive vocabulary (PV) and controlled active vocabulary (CAV), since in the studies referenced by Krashen (1989) only words’ recognition and spelling was tested. Therefore no FAV was examined and discussed by Krashen

(1989). It seems that the effect of comprehensible input on FAV acquisition is a neglected point in relevant research.

In recent years, some studies have been conducted to investigate the effect of different types of comprehensible input on “productive vocabulary” development. However, these studies need to be referenced cautiously, as the “productive vocabulary” in these studies are measured in different ways and may reflect different types of vocabulary knowledge. For example, in Pietila and Merikivi (2014) and Laufer and Nation (1999), Controlled Productive Vocabulary Test was adopted to assess students’ vocabulary size. In Hazrat and Hessamy (2013), researchers tested the target words by asking testees to fill in the blanks of given sentences. In each blank, the first letter of the target word was provided, so that the non-target words that fit the context can be avoided. Soleimani and Mahmoudabadi (2014) examined the impact of input and output tasks on “productive vocabulary”, but the “productive vocabulary” was measured by asking testees to make sentences with the given target words. Therefore, all of these studies in fact measure some other aspects of “productive vocabulary” rather than FAV.

Excluding these studies, a few studies are left that investigate the impact of input on FAV acquisition. Arash (2016) examined the effects of three types of input on productive vocabulary development. The three types include non-negotiated pre-modified input, negotiation of input without output, and negotiation plus output. To test “productive vocabulary”, testees were provided with images of the target words. Then they were asked questions on the images. The purpose of this activity was to invoke the testees to speak out the words. Whether the target words could be produced or not completely depended on the testee himself. Therefore, it was believed that the test in this study examined testees’ FAV. Arash (2016) found all the three

types of input facilitate FAV gains. Teng (2015) investigated impacts of extensive reading and extensive reading plus explicit vocabulary learning on “productive vocabulary” acquisition. In the test, the participants were asked to fill in the blanks of given sentences. It is believed that this test assessed FAV as well, as there was no hint of spelling or pronunciation of the target words in the test. Testees had to depend on themselves to retrieve and use the target words in the context. Results of the research indicated that both extensive reading and extensive reading plus explicit vocabulary learning considerably facilitated FAV development.

Both the two studies seem to prove that comprehensible input helps in FAV development. However, the documentary analysis in this sub-study seems to lead us to a different conclusion that comprehensible input is not facilitative, or at least not adequately facilitative to FAV development, as after large input for 12 months the FAV did not show significant increase. In other words, the sub-study tends to indicate that a large amount of comprehensible input fails to promote significant FAV growth. Different from FAV, the longitudinal sub-study in Chapter 4 shows that participants’ passive vocabulary (PV) and controlled active vocabulary (CAV) kept growing in the 12 months. This result to some extent backs up Krashen’s (1989) view, as the large amount of comprehensible input seems to facilitate the development of word recognition and spelling. Students’ FAV, which was not mentioned by Krashen (1989), seems not to have benefited from the large amount of comprehensible input.

This sub-study therefore yields results different from previous studies. A large amount of input did not result in a significant growth in FAV. Taken a closer look, it is found that the different outcomes may be attributed to participants’ different L2 proficiency levels. It should be noted that the L2 proficiency of lower intermediate, intermediate, higher/upper intermediate and advanced levels are often used arbitrarily

by researchers, and it seems that there is no set standards yet for proficiency evaluation. The arbitrariness sometimes causes confusions, and makes it difficult to have comparisons across studies. Therefore, it is necessary to examine participants' background information even if their L2 proficiencies have been tagged by researchers. Teng (2015) described his participants as "lower-intermediate" L2 learners. In his study, subjects were freshmen majoring in Business English. Their ages ranged from 18 to 20 years old. The mean score of the Vocabulary Size Test designed by Nation and Beglar (2007) was 21.36. The researcher therefore concluded that the subjects' English proficiency was lower-intermediate. Although subjects in this study did not take the same vocabulary size test as those in Teng (2015) did, it is postulated that they are of higher L2 proficiency level, as they are senior English-major students who have had concentrated professional English training for three to four years. That is to say, subjects in Teng (2015) may not have reached the stage where FAV threshold occurs yet, and that is possibly the reason why comprehensible input yielded positive results in Teng's (2015) study, but did not do so in this sub-study. It is therefore postulated that comprehensible input plays a positive role in FAV progress before FAV threshold occurs; when learners' L2 proficiency reaches a certain level and FAV stagnation occurs, even a large amount of comprehensible input does not help. In other words, comprehensible input may not be an effective factor to tackle FAV stagnation. Another study by Arash (2016) cannot be analysed and compared in this case, as no background information on the participants' proficiency is provided by the researcher.

It should be noted that the conclusion does not necessarily mean that input is not helpful at all to FAV acquisition. Although no significant FAV growth is detected in the 12-month longitudinal sub-study, no significant regression is detected either. The

large amount of input may do some work in preventing FAV from regressing, since frequent input can remind participants of the words by providing them with chances to encounter those words repeatedly. When FAV size reaches a certain level, a large amount of input may be helpful to keep FAV at the level, but not helpful enough to invoke further growth. Other factors may be needed to achieve this goal.

It seems that the output-training units are not substantially facilitative to FAV development as well. There are three output-training units in the course structure. As the longitudinal sub-study shows, no significant FAV gains occurred during the 12-month period. It seems that the output-training units did not work efficiently either. However, it may be hasty to conclude that just like input, output does not help in FAV growth. Compared with other units, the units for output training are in a salient disadvantage, as they only account for 20% of the whole course structure. Therefore, the insignificant role of output-training units in FAV development may be attributed to the inadequacy of them. As is shown in Chapter 6, pushed output was found by interviewees to be helpful to FAV development. Therefore, the limited number of output-training units may imply that more of them should be added to the current course structure.

It is only postulated that inadequate pushed output is one of the possible factors for FAV stagnation. The postulation needs to be further testified by empirical studies. If it is true, then the appropriate amount of pushed output needs to be investigated by empirical studies as well.

Thirdly, incidental acquisition seems not to be an efficient factor in FAV development. The interviews in Chapter 6 show that many teachers do not provide guidance in FAV development, and almost no teaching strategies have been adopted to specifically facilitate it. It can be concluded then that there is little intentional FAV teaching in

these units. If any FAV is obtained in classes, it is most likely acquired incidentally. However, the amount of FAV acquired incidentally may be rather limited, as no significant growth of FAV is detected in the 12-month longitudinal sub-study.

It seems that few previous studies have investigated incidental FAV acquisition. The only one that can be found was conducted by Yue, Dai, and Zhang (2012). Yue et al. (2012) compared effects of semantic processing and form processing on incidental vocabulary acquisition. In the study, the task of semantic processing was to translate the target words into L1, and then to match them with their L2 paraphrases; the task of form processing was to count the numbers of letters in the target words, and then to find letter combinations that match the given pronunciations. The study found that both the two types of processing significantly facilitated participants' incidental acquisition of PV and CAV. However, the study detected no considerable progress in FAV, which is a higher level of knowledge than PV and CAV. Both the study by Yue et al. (2012) and this sub-study indicate that the amount of FAV acquired incidentally is rather limited. And as is found in Chapter 6, attention to target words plays a positive role in FAV learning. Based on these findings, it seems rational to postulate that intentional learning may be more effective than incidental acquisition in FAV development. Therefore, more intentional learning should be conducted for FAV development. In other words, FAV should be developed more consciously and deliberately, rather than unconsciously and incidentally.

7.5 Conclusion

This sub-study answered the proposed questions. There are some factors in the course design that may help promote FAV progress. The first factor is pushed output. To invoke constant increase of FAV, more output-training units may need to be

integrated into the current course structure, or more activities that promote output may need to be integrated in teaching and learning. The second factor is intentional teaching and learning. Intentional teaching and learning seem to be necessary to promote FAV growth.

There are some factors that are detected not to be as powerful as expected. The first one is comprehensible input. With a large amount of input, no significant growth of FAV was detected. The second factor is incidental acquisition. The amount of FAV acquired incidentally seems to be limited. Therefore, comprehensible input and incidental acquisition may not be suitable tools to tackle FAV threshold phenomenon.

Chapter 8

Discussions, Findings, Implications and Recommendations

8.1 Introduction

In light of the previous chapters, this chapter examines whether the five research objectives have been addressed and achieved through the research. Based on the analysis of results, discussions were made, and five major findings pertaining to the research objectives were uncovered. Discussions are also made in this chapter to examine whether the research questions have been answered satisfactorily. It was found that some findings are supported by previous literature; some findings, by contrast, are different or even opposite to arguments in past studies. This chapter first presents findings pertaining to the research objectives and research questions. It then discusses the theoretical and pedagogical implications of the findings. Lastly, recommendations for free active vocabulary (FAV) development are provided.

8.2 Discussions

This section provides a general discussion on the results of this research. There are in total four sub-studies in this research, namely, the 12-month longitudinal sub-study, the word association sub-study, the interview and the documentary analysis. The four sub-studies work together to provide researchers with a more comprehensive view on L2 FAV development, in particular, the threshold phenomenon in FAV development. They also provide a more comprehensive view on the reasons for the occurrence of the threshold phenomenon, and the ways to tackle the problem.

The 12-month longitudinal sub-study shows that the Lex30 score did not change significantly over the 12-month period of time. According to the scoring standard of Lex30 test, only words beyond the first 1,000 frequent content words, high-frequency function words, proper nouns, and numbers would obtain points. Therefore, the result of Lex30 test indicates that, in the 12 months of the longitudinal sub-study students' FAV beyond these categories failed to grow. The test of writing in the longitudinal sub-study yields similar results. The Lexical Frequency Profile (LFP) by RANGE shows that in writing, words (no matter in terms of word types or word families) that are beyond the first 1,000 frequency level failed to increase significantly. Therefore, both the Lex30 test and the writing test indicate that FAV experienced stagnation in the 12-month longitudinal sub-study.

With regards to the relationship of FAV with passive vocabulary and controlled active vocabulary, the 12-month longitudinal sub-study shows that the participants' passive vocabulary and controlled active vocabulary grew significantly. In other words, the participants could recognise and spell more words in the 12-month L2 learning. The sub-study also shows that in both the first month and the twelfth month, the PV and

CAV are significantly correlated with each other. By contrast, FAV is not closely related with PV and CAV, as it is not significantly correlated with PV in the first and the twelfth month, and it is not significantly correlated with CAV in the twelfth month. It is only partially correlated with CAV in the first month. The results in this sub-study are consistent with the cross-sectional study by Laufer (1998). In Laufer (1998), there were two groups of participants. One group of participants were 26 16-year-old 10th graders in Israel who had been learning English as a foreign language for six years; the other group of participants were 22 17-year old students in the 11th grade in Israel, who had been studying English as a foreign language for seven years. Just like the participants in this sub-study, Laufer's (1998) participants primarily obtained English training from formal English classes. Laufer (1998) found that the two groups' FAV beyond the first 2,000 frequency level were not significantly different; by contrast, the PV and CAV of the 11th graders were significantly higher than those of the 10th graders. Both the study by Laufer (1998) and this sub-study tend to indicate that L2 learners' FAV tend to stagnate in its development, whereas PV and CAV tend to develop in a more linear way. In other words, when learners' English reaches intermediate, upper-intermediate and advanced proficiency levels, their capability to recognise and spell words does not help promote FAV growth significantly. The results of Laufer's (1998) study indicate that FAV development may involve a more complicated process, and the development of PA and CAV may not be adequate to promote FAV growth.

Then the associations of a group of frequently produced words (FPW) and a group of seldom produced words (SPW) were compared. The comparison shows that there are more semantic nodes in FPW than in SPW; the semantic nodes constitute the majority of links in the lexical network of FPW, but this is not the case in SPW. In terms of the

non-semantic nodes, the condition is different: the SPW contains significantly more non-semantic nodes than the FPW. Also, within the SPW, both the semantic nodes and non-semantic nodes account for about 50% in the lexical network. Within the semantic connections, it is found that the number of paradigmatic, syntagmatic, and encyclopaedia nodes are different. There are significantly more paradigmatic and syntagmatic connections in FPW than in SPW. In addition, syntagmatic connections account for a bigger proportion in FPW than in SPW, as it is 17.01% of semantic connections in FPW, but 9.36% in SPW. The encyclopedia nodes in FPW account for 9.13% in FPW, but its proportion in SPW is bigger, accounting for 18.71%. The paradigmatic connections are different as well. In FPW, there are more synonyms and antonyms than in SPW. By contrast, there are more syntactic nodes in FPW than in SPW. In terms of non-semantic nodes, the distribution in SPW and FPW is different as well. FPW has significantly less phonological-formal nodes and “no response” nodes than SPW had. The results of the word association sub-study may indicate that the organisation of the lexical network plays an important role in FAV development. The lexical network of FPW seems to be more semantised than SPW. Being more semantic may mean that the words can be accessed and retrieved more easily when the words are needed in expressions. In addition, the structure of the semantic connection in FPW seems to be more “appropriate” than that is in SPW, as semantic nodes that have shorter semantic distance to the target word tend to outnumber other semantic nodes.

The different features of the lexical network of FPW and SPW can shed some light on network construction to help promote FAV development. It seems that from SPW to FPW, the lexical network tend to get more and more semantic. As semantic nodes provide semantic paths in word access and retrieval, the tendency may be helpful for

words' access and retrieval when language users need it in context. In addition, the semantic connection in the lexical network tends to get stronger and tighter, as nodes which are more closely related to the target word grew significantly. The two tendencies are consistent with the Spreading-activation Model proposed by Collins and Loftus (1975), which assumes that lexical network develops in the direction that words are more easily accessed, activated, and retrieved. Based on the results of the word association sub-study in this research, it is assumed that some words may not get adequate semantic construction in the lexical network, and within the semantic connection the network is not adequately constructed to be stronger and tighter. Therefore, these words fail to become FPW.

In order to detect other possible reasons in addition to the cognitive reasons for the threshold phenomenon, interviews and documentary analysis were conducted. Interviews were also conducted to investigate effective learning and teaching strategies for FAV growth. The interview questions include students' consciousness of FAV in L2 study, learning strategies the learners adopted to enlarge FAV, teachers' guidance on FAV in L2 teaching, and teaching strategies that teachers adopted to help promote students' FAV development. When being asked whether they ever met difficulties in searching for suitable words to express the meaning they wanted to express, all the interviewees answered "Yes". They reported that they often had difficulties in searching for the most accurate word to use. In many cases they knew the word that they could retrieve was not the most accurate one, but they had to use the word, as they could not think of the most needed one. Thus some important information or subtle feelings may be lost in the context. The retrieval difficulty may find explanations from the word association sub-study. It was found in the word association sub-study that SPW has less semantic nodes than FPW, but more non-

semantic nodes than FPW. Therefore, when the target word is needed in context, the SPW is more difficult to be accessed and retrieved. As the difficulties in FAV usage are widely felt by the students, most of them reported that they knew the existence of this type of vocabulary knowledge, although they may not know the terminology of FAV.

The effective learning strategies to assist FAV development were also asked. The strategies reported by interviewees include making sentences, referring to a dictionary, taking notes on the word's meanings, analysing contexts, putting down example sentences, collecting synonyms, memorising authentic context for the target word. The effectiveness of these learning strategies are also consistent with the findings of the word association sub-study. It was found that all the strategies can help the semantic network construction of the target word in mental lexicon. Some of them also help the semantic network become stronger and shorter, such as the strategy of synonym collection. Constructing the lexical network in this direction may help words develop into FAV.

A group of questions were asked to investigate teachers' guidance on FAV learning and the effective teaching strategies that teachers adopted for FAV development. It seems that the guidance that learners could get from the teachers is rather limited, as in the four years' study, only one teacher told them that FAV is important. The student participants also reflected that very few suggestions were provided by their teachers; and almost no teaching strategies were adopted to promote FAV growth. This may be due to that fact that teachers are not conscious of the importance of FAV; or even if they are conscious of it, they lack adequate strategies to help learners with FAV development. However, it may not be appropriate to solely blame teachers for

this deficiency. As relevant research is far from adequate, teachers may not have enough resources to obtain guidance and information on this issue.

Some teaching strategies that are not adopted for FAV development but are thought to be effective are reported by interviewees. These strategies are providing synonyms of the target word to learners and negotiating the meanings of the target word with learners. The effectiveness of negotiation is supported by the empirical studies of Arash (2016) and Teng (2015). It is noticed that both synonym provision and negotiation can help construct and strengthen the semantic network of the target word, and thus they help make the word easier to be accessed, activated, and retrieved. The documentary analysis was adopted to investigate the possible reasons for FAV stagnation as well. By studying the course structure, it is found that units for output training were rather limited, as there were only 3 units for output training. These units are English Academic Thesis Writing, Undergraduate Thesis Writing, and Theory and Practice of Translation. Output practice is a significant part in these units. However, it is found that compared with other units, the output-training units are in a salient disadvantage. The output-training units only account for 20% in the course structure. According to the Output Hypothesis by Swain (1985, 1995), language production elicits language learning. Laufer (1998) and Henriksen (1999) pointed out that output can promote FAV acquisition; the word may remain passive and never be activated if learners are not pushed to produce the word. The empirical studies by Arash (2016) and Fuente and José (2002) show that pushed output plays a positive role in FAV development. This conclusion is consistent with what the interviewees reflected in the interview. In the interview the interviewees reported that they feel the output practice of the target word helps FAV development. From these results it is speculated that the

inadequacy of output-focused units is one of the factors for FAV stagnation, and more such units should be integrated into the current course structure.

8.3 Major Findings

8.3.1 Findings to Research Objective One: the Developmental Process of FAV

The first research objective of this study is to investigate the developmental process of FAV when L2 learners reach upper-intermediate to advanced L2 proficiency level. This objective was addressed by the 12-month longitudinal sub-study. The following questions in relation to the research objective were asked:

- What is the development pattern of FAV when L2 learners reach upper-intermediate to advanced proficiency level?
- If the threshold phenomenon occurs in the FAV development, what are the features of the threshold phenomenon?

The 12-month longitudinal sub-study was conducted to answer the two research questions. It tracked the development of FAV during a 12-month period, and found that the FAV did not grow in a linear way; instead, it stagnated during that period. It thus empirically supports the Active Vocabulary Threshold Hypothesis raised by Laufer (1991).

The research has answered the second research question as well. Several features were demonstrated in the FAV developmental process. First, the stagnation of FAV seems to be long-lasting, since the sub-study lasted for 12 months and saw no growth in FAV. Secondly, the stagnation of FAV seems to be “stubborn”, as it not only lasted for a long time, but also was not correlated with the growth of other types of vocabulary. Thirdly, the FAV threshold tends to occur widely, as in the sub-study it

emerged at most word frequency levels. The sub-study shows that words (no matter in terms of types or families) of the first 1,000 frequency level decreased significantly, whereas words of the second 1,000 frequency level, academic words and beyond did not increase significantly. This tendency is problematic, as according to Schmitt et al. (2001), the most frequent 2,000 words are usually for basic everyday communication, and words beyond 2,000 are more for authentic texts. Learners' LFP in this sub-study indicates that they tended to use much more basic words than NSs do. As pointed out by Cobb (2003), "over-use of basic vocabulary indicates, of course, under-use of other richer, more precise, and more varied vocabulary" (p. 399). This may account for the sense of vagueness felt by readers who have English as their native language in advanced L2 learners' writings (Cobb, 2003). Not only the NSs, the interview in this research shows that L2 learners themselves also feel their writing sometimes is "as plain as water". Therefore, even upper-intermediate to advanced L2 learners have a long way to go in FAV development.

In sum, the results of the longitudinal sub-study provide a comprehensive picture on FAV's developmental process in a 12-month period. During this period, words (no matter in terms of types or families) at most frequency levels experienced stagnation. Several features were captured from the sub-study as well, which are explained in Chapter 4.

8.3.2 Findings to Research Objective Two: the Relationship between FAV and Other Types of Vocabulary Knowledge

The second research objective is to investigate the relationship between FAV and passive vocabulary (PV), and the relationship between FAV and controlled active vocabulary (CAV). There are four research questions for the research objective:

- What is the developmental process of PV when L2 learners reach upper-intermediate to advanced L2 proficiency level?
- What is the relationship between FAV and PV?
- What is the developmental process of CAV when L2 learners reach upper-intermediate to advanced L2 proficiency level?
- What is the relationship between FAV and CAV?

The longitudinal sub-study has answered these research questions. It not only tracked the developmental process of FAV, but also tracked the developmental process of PV and CAV in the 12-month period of time. It turns out that during the period, both testees' PV and CAV grew significantly. With regard to PV, except words in the 2,000 level and in the 10,000 level, words of all other frequency levels increased significantly. These included the 3,000 level, the 5,000 level, and the academic words. The reason why words from the 2,000 frequency level did not have significant growth may be that the high-frequency words had already been grasped well by the upper-intermediate to advanced ESL learners, and there was almost no more space for further increase. The lack of growth for words in the 10,000 frequency level may be due to the fact that it is a very low frequency level. In other words, the chances to use words in the 10,000 frequency level are slim. Therefore, learners may not have many chances to encounter these words in their study. For this reason, scores in the 10,000 level did not change significantly and maintained low. With regard to CAV, words at all frequency levels developed significantly. This indicates that in the 12-month period, the testees not only could recognise more words, but also could spell more words. By contrast, FAV did not grow significantly in that period. It indicates that in upper-intermediate to advanced proficiency stage, FAV does not grow with the

growth of PV and CAV. The result is different from Laufer's (1998) speculation that a larger PV size may indicate a larger FAV size. This sub-study shows that this may not be the case, at least when L2 learners reach upper-intermediate to advanced proficiency level. The gap between PV and FAV, and between CAV and FAV in fact becomes larger.

8.3.3 Findings to Research Objective Three: the Lexical Network Construction in FAV Development

The third research objective is to explore how lexical network is constructed in the process of FAV development. As people's mental lexicon is like a "dark box", the procedure of network building cannot be observed directly. Therefore, the lexical organisations of two types of vocabulary are compared. One is vocabulary that is seldom produced, or highly unlikely to be produced by learners in writing and speaking; the other is vocabulary that is frequently produced, or is highly likely to be produced by learners in writing and speaking. It is assumed that if words develop from SPW to FPW, then the difference between the two in lexical organisation may shed some light on the network construction procedure in the process. Based on the research objective, one research question is proposed:

- What are the differences (if any) of the mental lexicon organisation between frequently produced words and seldom produced words?

The third research objective is addressed via word association tests. The word association tests found that there were a number of differences in the lexical organisation of the two types of vocabulary. First, the semantic and non-semantic distribution in the network is different. FPW turned out to have significantly more semantic nodes than SPW, and have significantly fewer non-semantic nodes than

SPW. In terms of the proportion of semantic and non-semantic nodes in the network, semantic nodes of FPW were dominant in the network, while non-semantic nodes only constitute a small part in the network. By contrast, in SPW, although semantic nodes outnumber non-semantic nodes, the advantage of semantic nodes is not as evident.

Secondly, the semantic distribution is different. The number of paradigmatic and syntagmatic nodes of FPW is significantly higher than that of SPW, but there is no significant difference in the number of encyclopaedia nodes between the two types of words. The ranking of semantic nodes in FPW is paradigmatic > syntagmatic > encyclopaedia; the ranking of semantic nodes in SPW is paradigmatic > encyclopaedia > syntagmatic.

Thirdly, the non-semantic distribution is different. FPW have less phonological-formal and “no response” nodes than SPW do. The percentages of phonological-formal nodes in non-semantic connections are different as well. The percentage tends to be higher in SPW than in FPW. Both the number and percentage of “no response” nodes tend to decrease in the process of SPW-to-FPW development.

Based on the results of this sub-study, the lexical construction procedure in FAV development can be portrayed. When a word can be recognised easily, but are highly unlikely to be produced, almost half of its nodes in its lexical network are non-semantic. These non-semantic nodes become a hindrance when L2 users try to access and retrieve the word from mental lexicon. The other half of the nodes are semantic nodes. Most of the semantic nodes at this stage are paradigmatic and encyclopaedic nodes. These paradigmatic and syntagmatic nodes can offer L2 users information on the target word in various perspectives, especially in synonyms, antonyms, social and cultural knowledge, grammar, and collocations. Therefore, the word not only can be

correctly recognised, but also can be well comprehended by the language user. At this stage, there are not as many syntagmatic nodes as paradigmatic and encyclopaedic nodes; therefore, the lexical network does not provide the learners with adequate paths to retrieve the target words' syntactic or compound/chunk information for production in context. With further development, semantic nodes tend to grow, and non-semantic nodes tend to shrink, until semantic nodes obtain salient advantage over non-semantic nodes and become dominant in the lexical network. At this stage, not only the size of semantic connections gets enlarged, the inner structure of semantic connections changes as well. Paradigmatic and syntagmatic nodes get significant growth, but encyclopaedia nodes do not. The result of this change is that the syntagmatic nodes surpass the encyclopaedia nodes and become the second biggest type of response. In paradigmatic connections, synonyms get prior development and become the majority in paradigmatic connections. Therefore, when target words are needed in oral or written expression, a large number of paradigmatic nodes (especially synonyms) and adequate encyclopaedia nodes would offer numerous channels for target words to be reached and retrieved. Meanwhile, a great deal of syntagmatic (especially syntactic) nodes are accessed and activated as well, making the target word ready to be produced in oral and written contexts.

In sum, the word association sub-study captured three possible features in FAV development. The first feature is that the lexical network tends to get semantised; the second feature is that different types of nodes tend to change at different speeds and in different directions (increasing or decreasing); the third feature is that the connections of the lexical network tend to get stronger and tighter. It is postulated that in L2 learning some words may not obtain enough network construction towards these directions; therefore, they fail to further develop into FAV.

8.3.4 Findings to Research Objective Four: Possible Factors Related to FAV Stagnation

The fourth research objective is to find possible factors in teaching and learning that may affect FAV development. Based on the research objective, two questions are designed:

- What are the possible factors in teaching that may affect FAV development?
- What are the possible factors in learning that may affect FAV development?

Findings from the interview and documentary analysis can help answer the two questions. On the teaching side, the first factor that may be related to FAV stagnation is that teachers seem not to attach great importance to FAV in teaching. Most teachers of the interviewees almost never mentioned FAV in class, let alone provided students with guidance on how to improve it. Therefore, realising the significance of FAV, students had to “grope” for effective learning methods by themselves. In most cases, the effectiveness of their trials depended on luck. Some strategies may turn out to be helpful, some may not; some may be helpful, but were only helpful to a specific person, not having a universal value.

The second factor in teaching is that little effort is made to help promote students’ FAV development. The interview reveals that few teaching strategies were adopted to facilitate students’ FAV development. However, it may be hasty to completely put the blame on teachers, as research in this field is rather limited, and it cannot give teachers sufficient guidance. Although researchers have detected the FAV threshold phenomenon in L2 study, such as Laufer (1991), Gu and Li (2013), Huang (2012), Lu (2008) and Y. Zhao (2011), there is a limited number of studies, especially empirical ones, to investigate how to tackle this problem. Even if teachers feel the great

necessity to help students enlarge FAV, they can find little reference for valuable suggestions. Therefore, relevant research is in dire need.

The third factor in teaching is detected from documentary analysis. It was observed that in the 12 months of longitudinal sub-study, only a few units were on L2 output. These units included English Academic Thesis Writing, Undergraduate Thesis Writing, and Theory and Practice of Translation. Compared with other units in the course structure, it seems that the number of units on output was in a salient disadvantage. Therefore, the limited number of units on output may be a factor for FAV stagnation, as in this condition L2 learners cannot get enough systematic output training in class.

Meanwhile, the course structure shows that the English-major students in the longitudinal sub-study received condensed English training, and the units covered a wide range of English-related topics, such as British and American literature, linguistics, intensive reading, etc. Therefore, it is postulated that students received a large amount of L2 input in that period of time. However, the FAV stagnation in the period indicates that the large amount of comprehensible input failed to promote significant FAV growth. In other words, when L2 learners have reached upper-intermediate to advanced L2 proficiency level, large comprehensible input may not be an effective way to tackle FAV stagnation.

It should be noted that the conclusion does not necessarily mean that input is not helpful at all in FAV acquisition. Although no significant FAV growth was detected in the 12-month longitudinal sub-study, no significant regression was detected, either. The large amount of input may do some work in preventing FAV from regressing, since frequent input can have students encounter the words repeatedly, and thereby it

has a reminding effect. When FAV size reaches a certain level, a large amount of input may help FAV maintain at the level, but may not help it have further growth.

On the learning side, several possible factors for FAV stagnation are detected as well. The first factor is insufficient effort. The interview shows that although students knew FAV, their effort to improve FAV seemed to be insufficient. The inadequacy can be reflected by two aspects. Firstly, the number of strategies they used is rather limited. Each interviewee only reported one or two strategies. Secondly, although some students endeavoured to improve FAV, they gave up later. This may be because without teachers' guidance, students tended to get discouraged or forget to keep doing it.

Insufficient intentional teaching and learning of FAV may be another factor for the stagnation. The interview shows that most teachers never reminded students to pay attention to FAV development, and almost no teaching strategies had been adopted to specifically facilitate this. Therefore, there is little intentional FAV teaching in class. On the students' side, learning strategies adopted for FAV learning are quite limited as well. Meanwhile, in the interview the students reflected that attention to target words plays a positive role in FAV learning. Based on these findings, it seems rational to postulate that more intentional teaching and learning is needed to promote FAV growth.

8.3.5 Findings to Research Objective Five: Effective Strategies to Tackle the Threshold Phenomenon

The fifth research objective is to find effective ways to tackle the FAV threshold phenomenon. The research question is:

- What are the effective strategies (if any) to tackle FAV threshold

phenomenon?

Some strategies were detected from the interview. On the learning side, the interview shows that making sentences with the target word, taking notes on the target word, putting down example sentences, and collecting synonyms of the word are effective strategies. On the teaching side, providing synonyms of the word to students, negotiating word meanings, analysing contexts, and encouraging students to make sentences with the target word could be helpful. Arousing students' attention to the target word is helpful as well, as it may elicit students to have intentional learning on the word. It is noticed that some teaching and learning strategies overlap with each other.

8.4 Implications

8.4.1 Theoretical Implications

The 12-month longitudinal sub-study in this research proved the existence of FAV threshold phenomenon. It also supports Laufer's (1991) view that FAV does not develop with PV and CAV. Moreover, the research has a further development on Laufer's (1991) FAV threshold hypothesis, as several features are captured in the phenomenon. First, when L2 learners are at upper-intermediate to advanced proficiency level, their FAV tends to stagnate at various word frequency levels. Except for words from the first 1,000 frequency level, words from the second 1,000 frequency level, academic words, and beyond all stagnate. Secondly, when L2 learners reach upper-intermediate to advanced proficiency level, FAV threshold tends to be long-lasting, since in the 12 months of sub-study FAV did not grow. Thirdly, FAV threshold seems to be "stubborn". This feature can be demonstrated by the fact

that through a concentrated L2 learning for 12 months, learners' PV and CAV got significant growth, but FAV did not.

In addition to the development on FAV threshold hypothesis, the research helps researchers and ESL teachers obtain a better understanding of L2 mental lexicon. As far as it is known, little research has attempted to explore learners' mental lexicon change in the process when words with high familiarity develop into FAV. Additionally, little is known on how lexical organisation interacts with memory search and retrieval in word production. It is indicated by this research that FAV development may involve readjustment in lexical organisation. There are three tendencies in the readjustment. The first tendency is that the lexical organisation tends to get semantised; the second tendency is that different types of nodes tend to change at different speeds and in different directions (increasing or decreasing); the third is that the lexical network connections tend to get stronger and tighter. The three tendencies show that in FAV development, the lexical network tends to develop in a way that the target word becomes more easily accessed, activated, and retrieved. This research therefore to some extent fills in the gap in literature.

8.4.2 Pedagogical Implications

Some pedagogical implications can be made from the research. First, teachers should be conscious of the existence of FAV threshold phenomenon. This may be the first step to tackle the stagnation problem. To raise the awareness, more research should be conducted in this area, so that teachers can get more chances to read studies on the phenomenon. Meanwhile, the issue should be integrated into textbooks and training courses for ESL teachers. More discussions on the topic are suggested to be conducted on the Internet, and more materials on L2 FAV learning should be posted

on the Internet, so that teachers can get more access to know the issue. Secondly, teachers not only need to know the phenomenon, but also need to have a better understanding of it. For example, teachers should know that FAV stagnation tends to occur widely at various word frequency levels. Teachers should also be aware that when L2 learners reach upper-intermediate to advanced proficiency level, a larger PV and CAV size may not substantially help the growth of FAV. In other words, when learners reach upper-intermediate to advanced proficiency level, promoting FAV growth by strengthening PV and CAV may not be an effective effort. Teachers should inform students of the phenomenon as well, including its features. Suggestions and guidance should be provided. For example, teachers are suggested to provide learners with effective learning strategies, such as collecting synonyms, making sentences with the target word, taking notes on the target word's meanings, discussing word meanings with others, and paying more attention to the target word. The software RANGE can be introduced to students, so that students can monitor their FAV in writing by using it. Test tools of Passive Vocabulary Levels Test, Controlled Active Vocabulary Test, and Lex30 test are suggested to be introduced to students as well, so that they can monitor their development of receptive vocabulary, controlled active vocabulary and free active vocabulary by doing these tests. On the learning side, the implication is similar. Students should be aware of the FAV threshold phenomenon, and have a better understanding of it.

Secondly, the study reveals that some strategies may be helpful to tackle the problem. As is discussed in Section 8.2.5, strategies such as making sentences with the target word, negotiating word meanings, collecting synonyms of the target word, and taking notes on the word's meanings may facilitate FAV development. In addition, persistent efforts should also be made, as FAV growth appears to be a long, slow process. With

regard to the course design, more units training learners' L2 output should be integrated into the courses. Meanwhile, more output-oriented activities should be conducted in classes.

The principles behind the effective strategies are suggested to be provided to students as well, so that they can have a better understanding of why these strategies work. These principles include: first, when words have been grasped with high familiarity, strategies should be adopted to facilitate the semantisation in mental lexicon construction; secondly, the network construction of synonym and syntactic nodes should be paid specific attention to.

8.5 Recommendations

Section 8.3.2 presents pedagogical implications obtained from the research. However, as reflected in the interview, these strategies seem to be far from enough. More teaching and learning tactics need to be adopted to promote FAV growth. Therefore, based on the principles concluded from this research, some tactics are recommended in this section.

8.5.1 Tactics to Strengthen Network Building

The first principle implied by the research is that teachers and learners should pay sufficient attention to network building in mental lexicon. When learners can recognise a target word with high familiarity, two directions should be followed in mental lexicon network building. The first one is semantisation. More semantic nodes should be integrated in the target word's network. To achieve this goal, it is recommended that traditional presentation style of the new words in textbooks be changed. Usually new words are listed in an alphabetic order in textbooks. In the list,

the words' pronunciations, lexical classes, and word meanings are provided as well. This presentation pattern offers little help in semantic network building, as it represents words in an isolated way. In order to overcome this weakness, it is suggested that new words are not presented individually. Instead, words should be presented in clusters, with other words that are semantically related to them. For example, in a lesson the new words for students to learn are "rice", "cook", "country", and "remote". Traditionally, the words would be listed in an alphabetic order as "cook", "country", "remote", and "rice". Student will have to encounter these words in a way that they are isolated with each other. However, if the words are clustered semantically, they will be presented in two groups. One is "cook" and "rice", and the other is "remote" and "country". This presentation will be more helpful in students' semantic network building of the words. In addition, not only word meanings should be provided in the presentation, various example sentences should be provided as well. For example, when the new word "cook" is presented, various examples of the word such as "cook the meal", "cooking class", "She likes cooking" can be provided. Word chunks or collocations, if they appear in the text, are also suggested to be given. For example, if "give" is a new word, and "given away" is used in the text, the collocation of "give away" and its example sentences should be presented in the vocabulary list as well. Moreover, target words should be designed to appear repeatedly in texts, as repetition is likely to be facilitative in FAV learning. Meanwhile, teachers should help students form good learning habits. When looking up a new word in textbooks or dictionaries, students should not only seek word meanings, but also check the word's semantic properties, such as its contexts, collocations, and synonyms. These tactics may facilitate the semantic network building.

Another suggestion for semantic network building is negotiation. As reflected in the interview, word negotiations by teachers are facilitative to FAV learning. Negotiations by students may be helpful as well, as the empirical study by Arash (2016) reveals that L2 learners can benefit from negotiations between themselves, or with teachers. Through word negotiations, such as clarification, confirmation or repetition of word meanings, more semantic nodes that are closely connected with the word will be integrated into the network. Therefore, in addition to word negotiations by teachers, students should be encouraged to have negotiations by themselves as well. Teachers can provide students with opportunities to conduct word negotiations on the target words. An activity, for example, is that students collaborate in groups to write a composition on a given topic. In the process, group members work together to collect proper words for the writing. Whether the selected words are appropriate or not will be discussed. Dictionaries and the Internet are encouraged to be used in the process. In the end, the teacher will join their discussion and help them with word choice.

The second direction in network building is that paradigmatic and syntagmatic nodes should be strengthened. Among these nodes the synonym and syntactic nodes should be strengthened in particular. Several activities are recommended in building synonym nodes. The first one is paraphrase. In paraphrase, students are pushed to collect words of similar meanings to explain the target words. Rewriting and retelling are helpful as well. In rewriting and retelling, students are asked to use other words to express the same meaning. Brainstorming can also be used. In brainstorming, students are asked to work together and collect synonyms of a target word as many as possible. With regard to building syntactic nodes, several tasks are recommended. First, after words can be recognised with high familiarity, adequate grammar input should be

provided by the teacher. Grammatical knowledge should be provided, and example sentences that can demonstrate the grammatical features of the target word should be given as well. Then various types of grammatical practice are encouraged to be carried out. For example, in the “matching” exercise, students would be asked to connect the target word with its correct usage. In “true or false”, students would be asked to judge whether the word is used correctly or not in a sentence. In “filling in the blanks”, students would be asked to use the target word with a correct syntactic form. In addition to input, output of grammar should be encouraged as well. This study recommends that students be asked to create sentences with the target words. If syntactic mistakes occur in the sentence, feedback should be offered in time, and mistakes should be corrected. With these measures, syntactic nodes may obtain more chances to be built in lexical network, and those syntactic nodes that have already existed may get strengthened and become closer to the target word in the network.

8.5.2 Tactics to Strengthen Intentional FAV Learning

The second principle implied by the research is that intentional learning may work more efficiently than incidental learning in FAV development. Therefore, more intentional learning should be integrated. It is recommended that more pushed output activities be conducted, as in pushed output, students are “forced” to produce the target words intentionally. In addition to sentence-making, gap-filling is another pushed-output activity. In the gap-filling exercise, proper context is provided to students, and they are required to fill in the gap with suitable words. After the exercise is completed, feedback and explanations from the teacher should be given. This tactic is supported by the findings of Arash (2016) that if words are pushed to be produced and then feedback is given from the teacher, students’ ability to produce the words

would progress to a great extent. It is believed that pushed-output activities not only help students set up paradigmatic and syntagmatic connections in mental lexicon, but also make students more confident to use the target word. With confidence students will not avoid using those words in writing and speaking.

Another recommendation is that a unit on FAV development be designed and integrated into the current course structure. With further and deeper research, the development pattern of FAV and the threshold phenomenon will be better understood by ESL researchers and educators. This will provide foundation to design a unit for FAV development. The unit will provide students with a systematic training of FAV. Multiple strategies will be adopted in class; words from the second 1,000 frequency level, academic words and beyond will be the target words to be trained, as words from these frequency levels tend to stagnate. As FAV development seems to be a slow, time-consuming process, the ideal condition would be that the unit be provided in every academic year as an optional unit for ESL learners with a good command of English to take.

The last suggestion in intentional FAV learning is that teachers adopt the tool RANGE to check students' LFP in writing. By comparing students' LFP with that of native speakers, students' word production capability in each word frequency level can be estimated, and the improvement will be better directed. Students should be taught to use RANGE by themselves as well, so that they can have self-management in FAV development. By applying the software, students can monitor their LFP in writing by themselves. In this way, student will be more conscious about their weaknesses in word production, and they will be more targeted in FAV learning. For example, if RANGE shows that someone's academic words in an argumentative

writing is very low, then the student can work more on the academic words and try to elevate the percentage of that part of words in his writing.

8.6 Conclusion

This chapter has provided an overall reflection of the findings in relation to the research objectives and questions in this study. Based on the findings, the research objectives have been achieved, and questions have been answered. The chapter then discusses the theoretical and pedagogical implications of the research, which is followed by recommendations on strategies that may help to tackle FAV threshold phenomenon.

The following chapter is a conclusion of the thesis. It will provide an overview of the entire research. There is a reflection of the research journey and achievements from the journey. Limitations of the research will also be disclosed, followed by possible directions for future research.

Chapter 9

Conclusion

9.1 Introduction

This chapter is a conclusion of the entire thesis. It gives an overview of the research, including a reflection on the research journey, a summary of the major findings, and a discussion on the significance of the research. Then it looks into the possible directions for future research.

9.2 The Research Journey

This research was conducted to fulfil the doctoral degree requirement of the University of Tasmania. As stated in the Introduction Chapter, the initial inspiration of the study was derived from the student researcher's own teaching experience. In her eight-year ESL teaching career, the researcher found that her English-major students, including those senior students who had been studying English as a second language for over ten years, were confined to the most common, unnuanced words in writing. The overuse of the most common, unnuanced words always made students' writing vague, and caused difficulties in expressing different moods and connotations. This problem triggered her interest in having a deeper investigation in L2 free active

vocabulary (FAV) acquisition. Then the student investigator did literature reading on the topic, which confirmed that the problem she encountered has also been experienced by others. It seems that there is a big gap between native speakers (NSs) and ESL learners in the FAV size. The literature reading further sparked her interest to explore this issue. Finally, the PhD study gave her an opportunity to formally conduct the exploration.

The research aimed to have a deeper understanding of FAV acquisition and to find possible ways to tackle the FAV threshold phenomenon if it occurred. It had five objectives. The first objective was to track the developmental process of FAV. The second objective was to investigate the relationship between FAV and passive vocabulary (PV), and the relationship between FAV and controlled active vocabulary (CAV). The third objective was to explore how lexical network was constructed in the process of FAV development. The fourth objective was to find possible factors in teaching and learning that may affect FAV development. The fifth objective was to find effective ways to tackle the FAV threshold phenomenon, if it did occur in the research.

To achieve the research objectives, several sub-studies were carried out. First, a 12-month longitudinal sub-study was conducted. In the first month, the participants were asked to complete a series of tests, including the Passive Vocabulary Levels Test, the Controlled Active Vocabulary Levels Test, and the Lex30 test for free active vocabulary. A composition of no less than 200 words was collected as well, to measure the testees' free active vocabulary in contexts. In the 12th month, the tests and composition were conducted again. Then results in the first month and the 12th month were compared. The purpose of the sub-study was to examine learners' FAV development pattern in the 12-month period of time. Another purpose of the sub-

study was to investigate the relationship between FAV and PV, and the relationship between FAV and CAV.

During the same period as the longitudinal study, the second sub-study was conducted as well. In this second sub-study two types of tests were conducted, one was Vocabulary Knowledge Scale test, and the other was word association test. The Vocabulary Knowledge Scale test was in a Likert-scale form, and it was adopted to measure testees' familiarity with the target words. The word association tests were adopted to examine the lexical organisation of the target words. There were two types of words, one was words that were highly likely to be produced by upper-intermediate to advanced L2 learners, and the other was words that were highly unlikely to be produced by them. Both these two types of words were high-frequency words used by native speakers. The Vocabulary Knowledge Scale test indicated that the testees were highly familiar with both of the two types of target words. Then the lexical organisations of the two types of words were compared. The purpose of the sub-study was to investigate the change of lexical organisation in FAV development.

The first and second sub-studies adopted quantitative approaches. The third and fourth sub-studies adopted qualitative approaches. After the longitudinal sub-study was completed, three senior students from the department were interviewed. The interview questions were designed from four aspects. The first aspect was on students' consciousness on FAV learning; the second aspect was on the learning strategies (if any) students adopted in FAV learning; the third aspect was on whether teachers have attached importance to FAV in teaching; and the fourth aspect was on effective teaching strategies (if any) to facilitate students' FAV development. Surrounding these aspects 14 questions were designed. When the interview was formally conducted, each interviewee was interviewed individually, so that they

would not be influenced by others. Each interview took around five to seven minutes. The purpose of the interview was to detect possible factors for FAV stagnation, and to detect effective strategies to promote FAV development.

Originally, the research involved the three independent sub-studies stated above. However, with further honing on the research plan, it was found that class instruction was a significant part in ESL learning, and should not be ignored in the investigation. Therefore, another sub-study was integrated into the original plan. The added sub-study was documentary analysis. In the sub-study, the course structure in the period of the 12-month longitudinal sub-study was collected. Then the units were examined and analysed. The Syllabus of English Courses for College English Majors in China was examined and analysed as well. The purpose of the sub-study was to find possible factors in course design that may affect FAV development.

In general, all the sub-studies were conducted smoothly, and all the research objectives have been achieved. Moreover, the entire research project was completed within a reasonable time frame. In terms of the data, as both quantitative and qualitative methods were adopted in the research, there was a triangulation of the data source. The research instruments were fully justified by prior studies and theories, and by the carefully designed pilot study. The careful selection of the participants ensured the generalizability of the research findings. The data analysis was conducted by using professional tools, including AntConc, RANGE and SPSS. In sum, every care has been made for the completion of the research project.

9.3 Summary of the Major Findings

This research uncovered five significant findings pertaining to the research objectives. A brief summary of the findings will be presented in this section.

First, it is found that the FAV threshold phenomenon exists when L2 learners reach upper-intermediate to advanced L2 proficiency level. Three features were captured in the threshold phenomenon. The first feature is that the threshold phenomenon is long-lasting; the second feature is that the threshold phenomenon is “stubborn”, as it does not grow with the growth of PV and CAV; the third feature is that the FAV threshold tends to occur widely, as it occurs at most of the word frequency levels.

Secondly, it is found that both PV and CAV keep growing when L2 learners reach upper-intermediate to advanced proficiency level; FAV is not significantly correlated with PV and CAV at this stage, and the gap between PV and FAV, and between CAV and FAV become larger.

Thirdly, there are three tendencies in lexical network construction when a word is developing into FAV. The first tendency is that the word’s lexical network tends to be semantised; the second tendency is that some types of nodes tend to increase in the process, while other types of nodes tend to decrease; and they change at different speeds. The third tendency is that the connections of the lexical network tend to get stronger and tighter. In summary, the lexical network tends to develop in a way that the word becomes easier to be accessed, activated, and retrieved.

Fourthly, it is found that there are possible factors for FAV stagnation in both ESL teaching and learning. In teaching, the first factor is that teachers seem not to have attached great importance to FAV in classroom instruction. The second factor is that little effort has been made to facilitate students in FAV development. The third factor is that there are inadequate units for output training in the curriculum. In learning, there are several possible factors for FAV stagnation as well. The first factor is that students’ effort to improve FAV may not be enough, as the strategies students adopted were rather limited; sometimes the effort was not persistent, as it was given

up later. The second factor is that there is insufficient intentional learning in FAV study.

Fifthly, some effective teaching and learning strategies are detected from the research. The learning strategies include sentence making with the target word, note-taking on the target word, putting down example sentences, collecting synonyms, etc. The effective teaching strategies include providing synonyms of the target word, negotiating word meanings, analysing the word's contexts, encouraging students to make sentences with the word, and raising students' attention to the word. In general, the proposed research objectives have been fulfilled, and the research questions have been answered. The implications of the research have been discussed in Chapter 8.

9.4 Significance of the Research

As pointed out in the chapter of Introduction, little prior research has adopted a long-time longitudinal study to track the developmental process of FAV, let alone to track the growth of FAV, PV and CAV simultaneously. In this research, a longitudinal sub-study lasting for 12 months was carried out, and the research unfolded the FAV threshold in the 12-month period of time. Therefore, the sub-study provides a vigorous evidence and a robust support for Laufer's (1998) FAV threshold hypothesis. Additionally, it uncovered more details on the phenomenon, which helps ESL teachers and researchers obtain a better understanding of it.

The research also explored the change of mental lexicon in FAV development. In addition to seeking external reasons in teaching and learning for the FAV threshold phenomenon, this research attempted to investigate the inner mechanism in FAV development. The word association sub-study investigated the lexical network of words that were highly likely to be produced by upper-intermediate to advanced L2

learners, and words that were highly unlikely to be produced by them. Then the lexical organisations of the two were compared. It is assumed that the difference between the two may shed some light on how the network in mental lexicon is constructed when words develop into FAV. Some possible tendencies in the process were discovered, and the sub-study may provide a kick-off for further research in this direction.

Another devotion of this research is that it broadens the research view on FAV study. The school courses offer a major source for L2 learners to receive ESL instruction, but it had been seldom integrated in FAV research. By observing the units learners took at the time of the longitudinal sub-study, some possible drawbacks in course design were detected. Lastly, by interviewing the learners, more first-hand data were obtained to explore students' attitude to FAV learning, their strategy use, and teachers' teaching on FAV.

In general, the research provides robust evidence for FAV threshold hypothesis. It also broadens the view to investigate FAV threshold phenomenon to mental lexicon and course design. Moreover, the research may bring a positive impact on FAV learning and teaching, as it raises more attention to the threshold issue, offers ESL teachers and learners more guidance in FAV development, and help them understand, predict, and tackle the problem more efficiently.

9.5 Possible Directions for Future Research

Although the research has achieved the proposed objectives and answered the research questions, it was found that there are still numerous unanswered questions on FAV threshold phenomenon. More studies are needed to have an even deeper insight into the phenomenon. Moreover, it is acknowledged that limitations exist in this

research. It is therefore hoped that future research can answer the questions and overcome the limitations.

Although a part of the research is a 12-month longitudinal sub-study, it is strongly felt that an even longer longitudinal study is in dire need. The outcome of this longitudinal sub-study reveals that in the 12 months of time learners' FAV did not grow. Then questions emerge: when does the threshold start to occur? Does the threshold occur periodically, or does it occur once and stay there permanently? In other words, is the threshold temporary and periodical, or is it eternal? The longitudinal sub-study in this research only followed upper-intermediate to advanced L2 learners for 12 months. A more ideal condition would be that the researcher follows a group of participants throughout the four-year English-major study at a university, and if possible, keeps following the participants who continue with their Master's and PhD study. This will provide a more comprehensive picture on FAV development.

Another issue that needs to be verified by future research is whether the effective teaching and learning strategies found in this research are really effective. These strategies include the ones detected from the research and the ones recommended based on the principles concluded from the research. In the research, the effectiveness of these strategies was based on the interview and the documentary analysis, but it was not evidenced by empirical studies. In other words, effective learning and teaching strategies not only should be guided by experience and theories, but also need to be tested by empirical studies. For example, based on the findings of this research, it is postulated that pushed output activities can facilitate FAV learning. However, whether the postulation is correct and how much pushed output will be adequate to promote FAV growth need to be testified by empirical studies. If these

pushed output activities are confirmed to be effective by empirical studies, they can then be adopted more frequently in FAV teaching and learning, and the threshold problem will be more effectively tackled in this way. Therefore, it is suggested that more empirical research be conducted to verify the effectiveness of these strategies. It is believed that with quantitative-evidence support, the conclusion will be more convincing, and more instructive to FAV learning.

Empirical studies will help answer a significant question raised by Laufer (1998) as well. The question is whether the threshold phenomenon occurring in FAV development reflects the nature of FAV learning, or whether it is just the consequence of certain types of teaching and learning. In other words, whether FAV stagnation will occur irrespective of the load and style of teaching and learning, or whether it is a result of them. The question seems to be significant, but it has not been satisfactorily answered yet until now. Longer longitudinal studies and studies testifying teaching and learning strategies will help answer the question.

It is also expected that the limitations of the present study can be overcome in future research. Due to time constraints, the present research only measured the breadth of FAV. The development pattern of the depth of FAV was not touched upon. Therefore, it is unknown yet whether the threshold phenomenon only occurs in the breadth of FAV, or in both the breadth and depth of it. In addition, due to the limited research conditions, the sample sizes in this research are comparatively small. As it is known, the sample size is one of the factors that influence the robustness of quantitative empirical studies. Therefore, it is expected that in future studies bigger sample sizes can be obtained.

Lastly, the word association approach adopted in the study can only be applied to observe the network structure of the target words in learners' mental lexicon. More

detailed information, such as the density of the lexical network, the strength and length of connections the target words have with other nodes in the network, cannot be measured by word association tests. Other tools, such as Event-related Potentials (ERP), eye-movement tests, and response-time tests, may need to be applied to obtain the information in future research.

9.6 Conclusion

Free active vocabulary is an important part in ESL learning. However, the threshold phenomenon seems to be a big obstacle in FAV development. This research project attempts to have a deeper and more comprehensive investigation into the phenomenon. It is expected that the research could help ESL teachers and learners obtain a better understanding of the threshold's occurrence, features, and inner mechanism. It is also expected that the research can provide some guidance in teaching and learning to tackle the threshold problem. Meanwhile, although the proposed questions have been well answered, the research has led to new queries that need to be explored and answered in future studies. Therefore, the research is not an end, but rather a commencement opening up more possibilities for future research.

The research has finally reached its finishing line. Although the student researcher encountered numerous challenges in the journey, she benefitted a lot from the journey as well. The researcher has gained a deeper insight in the FAV threshold phenomenon. She has also gained research experience, intellectual growth, and team spirit from the journey. Meanwhile, the researcher is in debt with all the people who helped her in the long journey. Without the cooperation of the participants, the guidance of the supervisors, and the assistance of staff at the University of Tasmania, it is impossible for the student investigator to complete this huge project. Therefore,

the student researcher would like to extend her dire gratitude to those who have given the generous and kind assistance throughout the research.

References

- Aitchison, J. (1994). *Words in the mind : An introduction to the mental lexicon*. Oxford: Blackwell.
- Albrechtsen, D., Haastrup, K., & Henriksen, B. (2008). *Vocabulary and writing in a first and second language: Processes and development*. New York: Springer.
- Arash, A. (2016). Effects of non-negotiated pre-modified input, negotiation of input without output, and negotiation of input plus pushed output on EFL learners' vocabulary learning. *Journal of Language Teaching and Research*, 7(4), 773-779. doi: 10.17507/jltr.0704.19
- Astika, G. G. (1993). Analytical assessments of foreign students' writing. *RELJ Journal: A Journal of Language Teaching and Research in Southeast Asia*, 24(1), 61-72.
- Bachoud-Lévi, A.-C., Dupoux, E., Cohen, L., & Mehler, J. (1998). Where is the length effect? A cross-linguistic study of speech production. *Journal of Memory and Language*, 39(3), 331-346.
- Barcroft, J. (2004). Effects of sentence writing in second language lexical acquisition. *Second Language Research*, 20(4), 303–334.

- Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research*, 22(4), 487-497.
- Bayazidi, A., & Saeb, F. (2017). Assessing reliability of two versions of vocabulary levels tests in Iranian context. *Advances in Language and Literary Studies*, 8(1), 30-43.
- Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17(6), 905-917.
- Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernbascher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). Sandiego, CA: Academic Press.
- Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28(3), 297-332. doi: 10.1016/0010-0277(88)90017-0
- Carolyn, Y., Ping, H., Gutchess, A. H., Hedden, T., Hiu-Ying Mary, C., Qicheng, J., & Yao, C. (2004). Category norms as a function of culture and age: Comparisons of item responses to 105 categories by American and Chinese Adults. *Psychology and Aging*(3), 379.
- Carroll, D. W. (1986). *Psychology of language*. Pacific Grove, California: Brooks/Cole Pub. Co.
- Channell, M. (1990). Vocabulary acquisition and the mental lexicon. In J. Tomaszczyk, B. Lewandowska-Tomaszczyk & R. Burchfield (Eds.), *Meaning and lexicography* (pp. 21-30). Amsterdam: Benjamins.
- Charmaz, K. (2006). *Constructing grounded theory : A practical guide through qualitative analysis*. London: Thousand Oaks : SAGE.

- Chen, S. (2006). Semantic representations of bilingual mental lexicon. *Foreign Language and Literature Studies*, 23(1), 6-11.
- Chen, S., Peng, J., Yang, H., Hou, L., & Fang, H. (2011). A study on the familiarity effects in representing English words in English-Chinese mental lexicon. *English Education in China*, 32(2), 1-16.
- Chen, X., & Zhang, J. (2008). Comparison of the cognitive factors in Chinese verbs and Chinese nouns. *Journal of Huanan Normal University (Social Science Edition)* (4), 119-123.
- Chow, I. H. S., Inn, A., & Szalay, L. B. (1987). Empirical study of the subjective meanings of culture between American and Chinese. *Asia Pacific Journal of Management*, 4(3), 144-151. doi: 10.1007/bf01732384
- Clark, E. (1995). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Coady, J., Magoto, J., Hubbard, P., Graney, J., & Mokhtari, K. (1993). High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 217-228). Norwood, NJ: Ablex.
- Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review*, 59(3), 393-424.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428. doi: 10.1037/0033-295x.82.6.407
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240-247.

- Collins, A. M., & Quillian, M. R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organisation of memory*. New York: Academic Press.
- Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308-325). Oxford: Oxford University Press.
- Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41(3), 365-397. doi: 10.1006/jmla.1999.2651
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50, 491-511. doi: 10.1016/j.jml.2004.02.002
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. doi: 10.2307/3587951
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684. doi: 10.1016/s0022-5371(72)80001-x
- Cramer, P. (1968). *Word association*. New York: Academic Press.
- Crutch, S. J., & Warrington, E. K. (2005). Gradients of semantic relatedness and their contrasting explanations in refractory access and storage semantic impairments. *Cognitive Neuropsychology*, 22(7), 851-876.
- Cui, Y. (2006). *An exploration into the development of lexical competence of Chinese tertiary English majors*. Doctoral dissertation, Shanghai Jiaotong University, Shanghai, China.

- Cui, Y. (2010). Word association tests and second language vocabulary acquisition: A Re-examination. *Contemporary Foreign Languages Studies*(7), 55-59.
- Cui, Y., & Wang, T. (2006). The development pattern and correlation of receptive vocabulary, productive vocabulary and depth of knowledge. *Modern Foreign Languages*, 29(4), 392 - 400.
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. *Elementary English*, 42, 895-901.
- de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 824-845.
- de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1001-1018.
- de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27-51). Amsterdam: John Benjamins.
- Deese, J. (1962). Form class and the determinants of association. *Journal of Verbal Learning and Verbal Behavior*(1), 79-84.
- Deese, J. (1966). *The structure of associations in language and thought*. Baltimore: Johns Hopkins Press.
- Dong, L., & Zhang, P. (2011). The non-semantisation of mental lexicon. *Journal of Southeast University (Philosophy and Social Science)*, 13(Supplement), 101-104.
- Duan, S., & Yan, C. (2004). Effects of multiple-choice annotation on ESL incidental vocabulary acquisition. *Foreign Language Teaching and Research*, 36(3), 213-218.

- EAT. (2015). *Edinburgh Association Thesaurus*. Retrieved from <http://www.eat.rl.ac.uk/>
- Education, C. M. o. (2000). *The syllabus of English courses for college English majors*. Beijing: Foreign Language Teaching and Research Press.
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, 21(2), 285-301.
- Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing*, 4(2), 139-155. doi: 10.1016/1060-3743(95)90004-7
- Entwisle, D. R. (1966). *Word associations of young children*. Baltimore: Johns Hopkins Press.
- Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *The American Journal of Psychology*, 74(3), 361-372. doi: 10.2307/1419742
- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal*, 31(2), 105-119. doi: 10.1177/003368820003100205
- Ferré, P., & Sánchez-Casas, R. (2014). Affective priming in a lexical decision task: Is there an effect of words' concreteness? *Psicológica*, 35(1), 117-138.
- File, K. A., & Adams, R. (2010). Should vocabulary instruction be integrated or isolated? *TESOL Quarterly*, 44(2), 222-249.
- Fitzpatrick, T. (2006). Habits and rabbits: Word associations and the L2 lexicon. *EUROSLA Yearbook*, 6, 121.

- Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *Vocabulary and Language Teaching*, 17(3), 319-331. doi: 10.1111/j.1473-4192.2007.00172.x
- Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, 27(4), 537-554.
- Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition*, 33(3), 373-398.
- Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *VIAL: Vigo International Journal of Applied Linguistics*(1), 55-73.
- Fitzpatrick, T., Wray, A., Playfoot, D., & Wright, M. J. (2015). Establishing the reliability of word association data for investigating individual and group differences. *Applied Linguistics*, 36(1), 23-50. doi: 10.1093/applin/amt020
- Fontana, A., & Frey, J. (2000). The interview: From structured questions to negotiated text. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., Vol. 2, pp. 645-672). Thousand Oaks: SAGE.
- Forster, K. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.
- French, R. M., & Jacquet, M. (2004). Understanding bilingual memory: Models and data. *Trends in Cognitive Sciences*, 8(2), 87-93.

- Fu, Y., Cui, Y., & Chen, H. (2009). A Longitudinal study of the developmental pattern of SL mental lexicon. *Foreign Language and Literature Studies*, 26(1), 16-23.
- Fuente, D. L., & José, M. (2002). Negotiation and oral acquisition of L2 vocabulary: The roles of input and output in the receptive and productive acquisition of words. *Studies in Second Language Acquisition*, 24(1), 81-112.
- Gina, N., Jennifer, L., Jill, C., Marco, A., & Guy, T. (2011). Examining multiple dimensions of word knowledge for content vocabulary understanding. *The Journal of Education*, 192(2/3), 49-61.
- Gong, R. (2008). The optimization of cognitive factors affecting the semantic network of L2 mental lexicon of college English learners. *Journal of PLA University of Foreign Languages*, 31(3), 56-62.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67-81.
- Gu, Q., & Li, X. (2013). Study on productive vocabulary size of non-English major students. *Journal of Tonghua Normal University*, 34(3), 80-83.
- Hazrat, M., & Hessamy, G. (2013). The impact of two types of vocabulary preparation on listening comprehension, vocabulary learning, and vocabulary learning strategy use. *Theory and Practice in Language Studies*, 3(8), 1453-1461. doi: 10.4304/tpls.3.8.1453-1461
- Heatley, A., Nation, P., & Coxhead, A. (2002). *Range* [Computer software]. Retrieved from www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317.

- Henriksen, B. (2008). Declarative lexical knowledge. In D. Albrechtsen, K. Haastrup & B. Henriksen (Eds.), *Vocabulary and writing in a first and second language* (pp. 22-66). New York: Springer.
- Hu, A., & Liu, S. (2013). Model of memory processing system in the retrieval of translated words. *Foreign Languages and Their Teaching*, 273(6), 49-52.
- Hu, Y. (2013). Effects of English and Chinese words' cultural difference on English reading comprehension. *Journal of Inner Mongolia Normal University(Educational Science)*, 26(11), 102-104.
- Huang, S. (2012). An evaluation of productive vocabulary capacity of English major students. *Journal of Beijing Second Foreign Langue University*, 212(12), 75-78.
- Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, 44(1), 2-31.
- Johnson, B., & Christensen, L. B. (2014). *Educational research : Quantitative, qualitative, and mixed approaches* (5th ed.). Thousand Oaks, California: SAGE.
- Jullian, P. (2000). Creating word-meaning awareness. *ELT Journal*, 54(1), 37-46.
- Källkvist, M. (1999). *Form-class and task-type effects in learner English: A study of advanced Swedish learners*. Lund: Lund University Press.
- Khazaeenezhad, B., & Alibabae, A. (2013). Investigating the role of L2 language proficiency in word association behavior of L2 learners: A case of Iranian EFL learners. *Theory and Practice in Language Studies*, 3(1), 108-115. doi: 10.4304/tpls.3.1.108-115

- Kittay, E. F., & Lehrer, A. (1992). Introduction. In A. Lehrer & E. F. Kittay (Eds.), *Frames, fields and contrasts* (pp. 1-20). Hillsdale, NJ: Lawrence Erlbaum.
- Kolers, P. A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior*, 2(4), 291-300.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464. doi: 10.2307/326879
- Krashen, S. (1998). Comprehensible output? *System*, 26(2), 175-182.
- Kvale, S., & Brinkmann, S. (2015). *InterViews : Learning the craft of qualitative research interviewing* (3rd ed.). Los Angeles: Sage Publications.
- La, H. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll, A. M. B. & D. Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289-307). Oxford: Oxford University Press.
- Lambert, W. E., & Tucker, G. R. (1972). *Bilingual education of children : the St. Lambert experiment*. Rowley, Mass: Newbury House Publishers.
- Laufer, B. (1989). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, 6, 10-20.
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75(4), 440-448. doi: 10.2307/329493
- Laufer, B. (1994). The lexical profile of second language writing: Does it change over time? *RELC Journal*, 25(2), 21-33. doi: 10.1177/003368829402500202

- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271. doi: 10.1093/applin/19.2.255
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436. doi: 10.1111/j.0023-8333.2004.00260.x
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33-51. doi: 10.1177/026553229901600103
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365-389.
- Leech, G. (1974). *Semantics*. London: Penguin Books Ltd.
- Leech, G. (1983). *Principles of pragmatics*. London: Longman.
- Leedy, P. D., & Ormrod, J. E. (2005). *Practical research : Planning and design* (8th ed.). Upper Saddle River, New Jersey: Merrill Prentice Hall.
- Lei, L., Wei, Y., Ye, L., & Zhang, M. (2007). A study of incidental vocabulary acquisition through writing by Chinese EFL learners. *Journal of PLA University of Foreign Languages*, 30(1), 53-56.
- Lextutor. (2017). *The Scoring Program for Controlled Active Vocabulary Levels Test*. Retrieved from <http://www.lexutor.ca/tests/levels/productive/>

- Li, Q. (2008). A study on the associative meaning and its embodiment in the color words of English and Chinese. *Journal of Southwest Agricultural University (Social Science Edition)* 6(3), 133-135.
- Li, R. (2003). Associative meaning in social and cultural perspective. *Journal of Xiangfan University*, 24(6), 76-78.
- Li, S. (2004). Cultural association of English words on colors. *Research of Administration Science*, 18(3), 110-111.
- Liu, S. (2001). Exploring word knowledge and its acquisition patterns: An experimental study of word meaning and affix. *Foreign Language Teaching and Research*, 33(6), 436-441.
- Liu, S., Fu, B., & Hu, A. (2012). Mental lexicon representation with L2 learners of different levels as shown in syntagmatic and paradigmatic semantic responses. *Journal of PLA University of Foreign Languages*, 35(2), 57-70.
- Lognostics. (2017). *The Lex30 Test Online*. Retrieved from <http://www.lognostics.co.uk/tools/Lex30/index.htm>
- Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48(1), 31-69. doi: 10.1111/1467-9922.00032
- Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology: Section A*, 53(1), 1-22.
- Lu, M. (2008). A study of productive vocabulary's development in breadth of knowledge. *Foreign Language Learning Theory and Practice*(2), 10-15.

- Maréchal, C. (1995). *The bilingual lexicon: Study of French and English word association responses of advanced learners of French*. Master's thesis, University of Dublin, Dublin, Ireland.
- McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.
- McNeill, D. (1966). A study of word association. *Journal of Verbal Learning and Verbal Behavior*, 5(6), 548-557.
- Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, 11(2), 29-38.
- Meara, P. (1984). The study of lexis in interlanguage. In A. David, A. Howart & C. Criper (Eds.), *Interlanguage* (pp. 225–235). Edinburgh: Edinburgh University Press.
- Meara, P. (1990). A note on passive vocabulary. *Second Language Research*, 6(2), 150-154.
- Meara, P., & Alcoy, J. C. O. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, 22(1), 222-236.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4(2), 142-154. doi: 10.1177/026553228700400202
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30. doi: 10.1016/s0346-251x(99)00058-5
- Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language*, 40, 25-40. doi: 10.1006/jmla.1998.2602

- Namei, S. (2004). Bilingual lexical development: A Persian-Swedish word association study. *International Journal of Applied Linguistics*, 14(3), 363-388.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12-25.
- Nation, P. (1984). *Vocabulary lists : Words, affixes, and stems*. Wellington: English Language Institute, Victoria University of Wellington.
- Nation, P. (1990). *Teaching and learning vocabulary*. New York: Heinle & Heinle Publishers.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Nation, P., & Heatley, A. (1994). *Range: A Program for the Analysis of Vocabulary in Texts* [Computer software]. Retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>
- Navracsics, J. (2007). Word classes and the bilingual mental lexicon. In Z. Lengyel & J. Navracsics (Eds.), *Lexical processes: Applied linguistics perspective. Clevedon: Multilingual Matters* (pp. 17-35). Toronto: Multilingual Matters Limited.
- Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, 10(2), 430-434.
- Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language*, 31(2), 237-260. doi: 10.1016/0749-596x(92)90013-n

- Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin*, 84(1), 93-116. doi: 10.1037/0033-2909.84.1.93
- Neuman, W. L. (2004a). *Basics of social research: Qualitative and quantitative approaches*. Boston: Pearson Education.
- Neuman, W. L. (2004b). *Social research methods: Qualitative and quantitative approaches*. Boston: Pearson Education.
- New, B., Ludovic, F., Christophe, P., & Marc, B. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review*, 13(1), 45- 52.
- Nissen, H. B., & Henriksen, B. (2006). Word class influence on word association test results. *International Journal of Applied Linguistics*, 16(3), 389-407. doi: 10.1111/j.1473-4192.2006.00124.x
- O'gorman, E. (1996). An investigation of the mental lexicon of second language learners. *Teanga: The Irish Yearbook of Applied Linguistics*, 16, 15-31.
- Palermo, D. S. (1971). Characteristics of word association responses obtained from children in grades one through four. *Developmental Psychology*, 5(1), 118-123. doi: 10.1037/h0031075
- Pietila, P., & Merikivi, R. (2014). The impact of free-time reading on foreign language vocabulary development. *Journal of Language Teaching and Research*, 5(1), 28-29. doi: 10.4304/jltr.5.1.28-36
- Piper, T. H., & Leicester, P. F. (1980). *Word association behavior as an indicator of English language proficiency*. Vancouver, BC: The University of British Columbia.

- Postman, L. J., & Keppel, G. (1970). *Norms of word association*. New York: Academic Press.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56(2), 282-308.
doi: 10.3138/cmlr.56.2.282
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19(2), 12-25.
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371.
- Read, J. (2000). *Assessing vocabulary*. New York: Cambridge University Press.
- Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209-227). Amsterdam, Netherlands: Benjamins.
- Renmin University of China. (2017). *Report on the Passing Rate of TEM4*. Retrieved from <http://sfl.ruc.edu.cn/displaynews.php?id=1741>
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.
- Richards, J. C. (2001). *The context of language teaching*. Beijing: Foreign Language Teaching and Research Press.
- Riegel, K. F., & Zivian, I. W. M. (1972). Study of inter- and intralingual associations in English and German. *Language Learning*, 22(1), 51-63.
- Roelofs, A. (1998). Lemma selection without inhibition of languages in bilingual speakers. *Bilingualism: Language and Cognition*, 1(2), 94-95.

- Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks: Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, 61(2), 292-323.
- Schmitt, N. (1998). Quantifying word association responses: What is native-like? *System*, 26(3), 389-401.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19(1), 17-36.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-88.
- Schwanenflugel, P. (1991). Why are abstract concepts hard to understand? . In P. Schwanenflugel (Ed.), *The psychology of word meaning* (pp. 223-250). Mahwah, NJ: Erlbaum.
- Seidman, I. (1998). *Interviewing as qualitative research: A guide for researchs in education and social sciences* (2nd ed.). New York: Teachers College Press.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Siok, H. L., & James, M. (2006). From receptive to productive: Improving ESL learners' use of vocabulary in a postreading composition task. *TESOL Quarterly*, 40(2), 295. doi: 10.2307/40264524
- Söderman, T. (1993). Word associations of foreign language learners and native speakers: The phenomenon of a shift in response type and its relevance for lexical development. In H. Ringom & T. Björkfors (Eds.), *Near-native proficiency in English* (pp. 91-182). Turku, Finland: Åbo Akademi University.

- Sökmen, A. (1993). Word association results: A window to the lexicons of ESL students. *JALT Journal* 15(2), 135-150.
- Soleimani, H., & Mahmoudabadi, Z. (2014). The Impact of Interactive Output Tasks on Developing Vocabulary Knowledge of Iranian EFL Learners. *Iranian Journal of Applied Linguistics*, 17(2), 93-113.
- Son, J.-S., Do, V. B., Kim, K.-O., Cho, M. S., Suwonsichon, T., & Valentin, D. (2014). Understanding the effect of culture on food representations using word associations: The case of “rice” and “good rice”. *Food Quality and Preference*, 31, 38-48. doi: 10.1016/j.foodqual.2013.07.001
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1), 41-78.
- Stolz, W. S., & Tiffany, J. (1972). The production of "child-like" word associations by adults to unfamiliar adjectives. *Journal of Verbal Learning and Verbal Behavior*, 11(1), 38-46. doi: 10.1016/s0022-5371(72)80057-4
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury.
- Swain, M. (1995). Three function of output in second language learning: Principles and practice in applied linguistics. Oxford: Oxford University Press.
- Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford: Oxford University Press.
- Tallerman, M. a. (2015). *Understanding syntax*. London: Routledge.

- Tan, X. (2006). A study of Chinese English learners' productive vocabulary development. *Foreign Language Teaching and Research*, 38(3), 202-207.
- Tan, X. (2007). A study of English learners' development of In-depth knowledge on productive vocabulary. *Foreign Language Education*, 28(2), 52-56.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology : Combining qualitative and quantitative approaches*. United States: SAGE.
- Teijlingen, V., Edwin, & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, 16(40), 33-36.
- Teng, F. (2015). The effectiveness of extensive reading on EFL learners' vocabulary learning: Incidental versus intentional learning. *Brazilian English Language Teaching Journal*, 6(1), 82-96.
- Theanalysisfactor. (2017). *How to Use Chi-Square Test*. Retrieved from <http://www.theanalysisfactor.com/observed-values-less-than-5-in-a-chi-square-test-no-biggie/>
- Thoughtco. (2017). *Function Word (Grammar)*. Retrieved from <http://www.thoughtco.com/function-word-grammar-1690876>
- Tong, S. (2009). An empirical study on the development of English major's oral productive vocabulary. *Foreign Language Research*, 150(5), 161-164.
- Tongji University of China. (2017). *Report on the Passing Rate of TEM8*. Retrieved from <http://sfl.tongji.edu.cn/content.dep?m=news&id=14648273962813872>
- Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12(4), 242-248.
- van Hell, J., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3), 193-211.

- van Hell, J., & Tanner, D. (2012). Second language proficiency and cross language lexical activation. *Language Learning*, 62(s2), 148-171. doi: 10.1111/j.1467-9922.2012.00710.x
- Von, H., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *Journal of Experimental Child Psychology*, 113(4), 569-586. doi: 10.1016/j.jecp.2012.08.001
- Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, 9(2), 172-185.
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(01), 33-52.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65.
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79-95.
- Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, 40(3), 360-376. doi: 10.1177/0033688209343854
- Wen, Q. (2006). A longitudinal study on the changes in speaking vocabulary by English majors in China. *Foreign Language Teaching and Research*, 38(3), 185-195.
- Wen, Q., Wang, L., & Liang, M. (2009). *Spoken and written English corpus of Chinese learners (2.0)*. Beijing: Foreign Language Teaching and Research Press.

- Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40.
- West, M. (1953). *A general service list of English words*. London: Longmans, Green.
- Wilks, C., & Meara, P. (2002). Untangling word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research*, 18(4), 303-324.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, 23(1), 41-69. doi: 10.1017/s0272263101001024
- Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, 30(3), 315-329. doi: 10.1016/s0346-251x(02)00017-9
- Wu, J., Lang, J., & Dang, Q. (2007). Incidental vocabulary acquisition and Task-induced Involvement. *Foreign Language Teaching and Research*, 39(5), 360-365.
- Wu, X., & Chen, X. (2000). Development of lexical competence in the EFL classroom setting. *Modern Foreign Languages*, 23(4), 349-360.
- Xue, G. Y., & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215-229.
- Yamamoto, Y. (2011). Bridging the gap between receptive and productive vocabulary size through extensive reading. *Reading Matrix: An International Online Journal*, 11(3), 226-242.
- Yang, S. (2007). The development of productive vocabulary in Chinese EFL learners. *Foreign Language and Literature*, 94(4), 254-259.

- Yu, W. (2011). Vocabulary problems and solutions in ESL writing. *Shandong Foreign Language Teaching Journal*, 140(1), 61-66.
- Yue, Y. (2006). "Acquiring" or "learning" English? *Foreign Language Teaching and Research in Basic Education*, 66(8), 21-23.
- Yue, Y. (2008). Research on L2 incidental vocabulary acquisition in China: Review and inspiration. *Journal of Taiyuan Normal University (Social Science Edition)*, 7(2), 165-167.
- Yue, Y., Dai, J., & Zhang, H. (2012). Effects of form and semantic processing on incidental vocabulary acquisition. *Journal of PLA University of Foreign Languages*, 35(2), 61-65.
- Zareva, A. (2007). Structure of the second language mental lexicon: How does it compare to native speakers' lexical organisation? *Second Language Research*, 23(2), 123-153.
- Zareva, A., & Brent, W. (2012). The 'promise' of three methods of word association analysis to L2 lexical research. *Second Language Research*, 28(1), 41-67.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567-595.
- Zhang, B. (2008). Different types of word links in the mental lexicon. *College Student Journal*, 42(2), 431-439.
- Zhang, G., Han, S., & Zhu, M. (2005). Development of L2 passive and active vocabularies: A comparative study on English and non-English majors. *Modern Foreign Languages*, 28(4), 374-382.
- Zhang, P. (2009). Word association and mental lexicon: A review of lexical network structure research. *Foreign Language Teaching and Research*(3), 71-82.

- Zhang, P. (2010a). Choice of prompt words and classification of responses in word association tests. *Journal of PLA University of Foreign Languages*, 33(1), 41-45.
- Zhang, P. (2010b). A comparative study of word association patterns in Chinese EFL learners' mental lexicon. *Foreign Language Teaching and Research*, 42(1), 9-16.
- Zhang, P. (2011). A comparative study of concreteness effect in the word associations of English and Chinese mental lexicon. *Foreign Language Learning Theory and Practice*(3), 54-62.
- Zhang, P., & Wang, Q. (2012). Effects of familiarity and frequency: Dynamic L2 mental lexicon. *Education Forum*(5), 106-107.
- Zhang, S. (2003). Study on vocabulary acquisition from the response type perspective. *Foreign Language Teaching and Research*, 35(4), 275-281.
- Zhang, S. (2004). Restructuring the L2 mental lexicon. *Journal of Sichuan International Studies University*, 20(2), 66-75.
- Zhang, S. (2005). Investigating connections in L2 mental lexicon with word association tests. *Journal of PLA University of Foreign Languages*, 28(2), 52-55.
- Zhang, S. (2006). Exploring the English L2 mental lexicon via word associations. *Modern Foreign Languages*, 29(2), 164-171.
- Zhang, S. (2008). The developmental course of the L2 mental lexicon: A longitudinal study. *Journal of Sichuan International Studies University*, 24(6), 120-124.
- Zhang, W., & Zhang, Q. (2007). Word retrieval mechanism and mental lexicon representation by advanced bilinguals. *Foreign Language Education*, 28(6), 50-53.

- Zhang, Y., & Zhang, J. (2014). Effects of word length on Chinese language production. *Journal of Psychology*, 46(9), 1232-1241.
- Zhao, P. (2013). Review and future directions of L2 mental lexicon research. *Journal of Chongqing University of Technology(Social Science)*, 27(6), 81-85.
- Zhao, Y. (2011). Fossilization in the output of high-frequency words from the “noticing” perspective. *Foreign Language Research*, 160(3), 100-102.
- Zhuang, J., & Zhou, X. (2001). Effects of time span on language production. *Journal of Psychology* 33(3), 214- 218.

Appendix A

The Passive Vocabulary Levels Test

右边竖行的解释适合左边竖行的哪个词？请在空格上填上相应的数字。请不要翻查字典，此成绩不会影响你的任何期末成绩。做题方法：每组先看右边竖行带空格的那三个词义，然后再从左面竖行的六个词中选择符合词义的词。不认识的字不要猜，空下就可。

Part A

1 copy	
2 event	_____ end or highest point
3 motor	_____ this moves a car
4 pity	_____ thing made to be like another
5 profit	
6 tip	

1 admire	
2 complain	_____ make wider or longer
3 fix	_____ bring in for the first time
4 hire	_____ have a high opinion of someone
5 introduce	
6 stretch	

1 accident	
2 debt	_____ loud deep sound
3 fortune	_____ something you must pay
4 pride	_____ having a high opinion of yourself
5 roar	
6 thread	

1 arrange	
2 develop	_____ grow
3 lean	_____ put in order
4 owe	_____ like more than something else
5 prefer	
6 seize	

1 coffee	
2 disease	_____ money for work
3 justice	_____ a piece of clothing
4 skirt	_____ using the law in the right way
5 stage	
6 wage	

1 blame	
2 elect	_____ make
3 jump	_____ choose by voting
4 manufacture	_____ become like water
5 melt	
6 threaten	

1 clerk	
2 frame	_____ a drink
3 noise	_____ office worker
4 respect	_____ unwanted sound
5 theater	
6 wine	

1 ancient	
2 curious	_____ not easy
3 difficult	_____ very old
4 entire	_____ related to God
5 holy	
6 social	

1 dozen	
2 empire	_____ chance
3 gift	_____ twelve
4 opportunity	_____ money paid to the government
5 relief	
6 tax	

1 bitter	
2 independent	_____ beautiful
3 lovely	_____ small
4 merry	_____ liked by many people
5 popular	
6 slight	

Part B

1 bull
2 champion
3 dignity
4 hell
5 museum
6 solution

_____ formal and serious manner
_____ winner of a sporting event
_____ building where valuable objects are shown

1 abandon
2 dwell
3 oblige
4 pursue
5 quote
6 resolve

_____ live in a place
_____ follow in order to catch
_____ leave something permanently

1 blanket
2 contest
3 generation
4 merit
5 plot
6 vacation

_____ holiday
_____ good quality
_____ wool covering used on beds

1 assemble
2 attach
3 peer
4 quit
5 scream
6 toss

_____ look closely
_____ stop doing something
_____ cry out loudly in fear

1 comment
2 gown
3 import
4 nerve
5 pasture
6 tradition

_____ long formal dress
_____ goods from a foreign country
_____ part of the body which carries feeling

1 drift
2 endure
3 grasp
4 knit
5 register
6 tumble

_____ suffer patiently
_____ join wool threads together
_____ hold firmly with your hands

1 administration
2 angel
3 frost
4 herd
5 fort
6 pond

_____ group of animals
_____ spirit who serves God
_____ managing business and affairs

1 brilliant
2 distinct
3 magic
4 naked
5 slender
6 stable

_____ thin
_____ steady
_____ without clothes

1 atmosphere
2 counsel
3 factor
4 hen
5 lawn
6 muscle

_____ advice
_____ a place covered with grass
_____ female chicken

1 aware
2 blank
3 desperate
4 normal
5 striking
6 supreme

_____ usual
_____ best or most important
_____ knowing what is happening

Part C

1 analysis
2 curb
3 gravel
4 mortgage
5 scar
6 zeal

_____ eagerness
_____ loan to buy a house
_____ small stones mixed with sand

1 contemplate
2 extract
3 gamble
4 launch
5 provoke
6 revive

_____ think about deeply
_____ bring back to health
_____ make someone angry

1 cavalry
2 eve
3 ham
4 mound
5 steak
6 switch

_____ small hill
_____ day or night before a holiday
_____ soldiers who fight from horses

1 demonstrate
2 embarrass
3 heave
4 obscure
5 relax
6 shatter

_____ have a rest
_____ break suddenly into small pieces
_____ make someone feel shy or nervous

1 circus
2 jungle
3 nomination
4 sermon
5 stool
6 trumpet

_____ musical instrument
_____ seat without a back or arms
_____ speech given by a priest in a church

1 correspond
2 embroider
3 lurk
4 penetrate
5 prescribe
6 resent

_____ exchange letters
_____ hide and wait for someone
_____ feel angry about something

1 artillery
2 creed
3 hydrogen
4 maple
5 pork
6 streak

_____ a kind of tree
_____ system of belief
_____ large gun on wheels

1 decent
2 frail
3 harsh
4 incredible
5 municipal
6 specific

_____ weak
_____ concerning a city
_____ difficult to believe

1 chart
2 forge
3 mansion
4 outfit
5 sample
6 volunteer

_____ map
_____ large beautiful house
_____ place where metals are made and shaped

1 adequate
2 internal
3 mature
4 profound
5 solitary
6 tragic

_____ enough
_____ fully grown
_____ alone away from other things

Part D

1 alabaster
2 chandelier _____ small barrel
3 dogma _____ soft white stone
4 keg _____ tool for shaping wood
5 rasp
6 tentacle

1 benevolence
2 convoy _____ kindness
3 lien _____ set of musical notes
4 octave _____ speed control for an engine
5 stint
6 throttle

1 bourgeois
2 brocade _____ middle class people
3 consonant _____ row or level of something
4 prelude _____ cloth with a pattern of gold or silver threads
5 stupor
6 tier

1 alcove
2 impetus _____ priest
3 maggot _____ release from prison early
4 parole _____ medicine to put on wounds
5 salve
6 vicar

1 alkali
2 banter _____ light joking talk
3 coop _____ a rank of British nobility
4 mosaic _____ picture made of small pieces of glass or stone
5 stealth
6 viscount

1 dissipate
2 flaunt _____ steal
3 impede _____ scatter or vanish
4 loot _____ twist the body about
5 squirm _____ uncomfortably
6 vie

1 contaminate
2 cringe _____ write carelessly
3 immerse _____ move back because of fear
4 peek _____ put something under water
5 relay
6 scrawl

1 blurt
2 dabble _____ walk in a proud way
3 dent _____ kill by squeezing someone's throat
4 pacify _____ say suddenly without thinking
5 strangle
6 swagger

1 illicit
2 lewd _____ immense
3 mammoth _____ against the law
4 slick _____ wanting revenge
5 temporal
6 vindictive

1 indolent
2 nocturnal _____ lazy
3 obsolete _____ no longer used
4 torrid _____ clever and tricky
5 translucent
6 wily

Part E

1 area			1 alter	
2 contract	_____	written agreement	2 coincide	_____ change
3 definition	_____	way of doing something	3 deny	_____ say something is not true
4 evidence	_____	reason for believing	4 devote and exactly	_____ describe clearly
5 method		something is or is not true	5 release	
6 role			6 specify	
1 debate			1 correspond	
2 exposure	_____	plan	2 diminish	_____ keep
3 integration	_____	choice	3 emerge	_____ match or be in agreement with
4 option	_____	joining something into a whole	4 highlight	
5 scheme			5 invoke	_____ give special attention to something
6 stability			6 retain	
1 access			1 bond	
2 gender	_____	male or female	2 channel	_____ make smaller
3 implementation	_____	study of the mind	3 estimate	_____ guess the number or size of something
4 license	_____	entrance or way in	4 identify	
5 orientation			5 mediate	_____ recognizing and naming a person or thing
6 psychology			6 minimize	
1 accumulation			1 explicit	
2 edition	_____	collecting things over time	2 final	_____ last
3 guarantee	_____	promise to repair a broken	3 negative	_____ stiff
4 media	_____	product	4 professional	_____ meaning 'no' or 'not'
5 motivation	_____	feeling a strong reason or	5 rigid	
6 phenomenon	_____	need to do something	6 sole	
1 adult			1 abstract	
2 exploitation	_____	end	2 adjacent	_____ next to
3 infrastructure	_____	machine used to move	3 controversial	_____ added to
4 schedule	_____	people or goods	4 global	_____ concerning the whole world
5 termination	_____	list of things to do at	5 neutral	
6 vehicle	_____	certain times	6 supplementary	

Appendix B

The Controlled Active Vocabulary Levels Test

填写词汇（每个词的前几个字母已给出），请注意时态、人称、数量等语法规则。为了避免对结论造成误导，请不要翻查字典。不认识的词请不要猜，空下就可。

Part A

1. I'm glad we had this opp_____（机会）to talk.
2. There are a doz_____（一打）eggs in the basket.
3. Every working person must pay income t_____（税）.
4. The pirates buried the trea_____（财宝）on a desert island.
5. Her beauty and ch_____（魅力）had a powerful effect on men.
6. La_____（缺乏）of rain led to a shortage of water in the city.
7. He takes cr_____（奶油）and sugar in his coffee.
8. The rich man died and left all his we_____（财富）to his son.
9. Pup_____（小学生们）must hand in their papers by the end of the week.
10. This sweater is too tight. It needs to be stret_____（撑大）.
11. Ann intro_____（介绍）her boyfriencd to her mother.
12. Teenagers often adm_____（尊敬,崇拜）and worship pop singers.
13. If you blow up that balloon any more it will bu_____（破裂）.
14. In order to be accepted into the university, he had to impr_____（改善）his grades.
15. The telegram was deli_____（传递, 投递）two hours after it had been sent.
16. The differences were so sl_____（轻微）that they went unnoticed.
17. The dress you're wearing is lov_____（可爱）.
18. He wasn't very popu_____（受欢迎的）when he was a teenager, but he has many friends now.

Part B

1. He has a successful car_____（职业, 事业）as a lawyer.
2. The thieves threw ac_____（酸性物质）in his face and made him blind.
3. To improve the country's economy, the government decided on economic ref_____（改革）.
4. She wore a beautiful green go_____（晚礼服, 裙子）to the ball.
5. The government tried to protect the country's industry by reducing the imp_____（进口）of cheap goods.
6. The children's games were amusing at first, but finally got on the parents' ner_____（神经）.
7. The lawyer gave some wise coun_____（忠告, 劝告）to his client.
8. Many people in England mow the la_____（草坪）of their houses on Sunday

morning.

9. The farmer sells the eggs that his he_____ (母鸡) lays.
10. Sudden noises at night sca_____ (使.....害怕) me a lot.
11. France was proc_____ (宣布) a republic in the 18th century.
12. Many people are inj_____ (受伤) in road accidents every year.
13. Suddenly he was thru_____ (用力推) into the dark room.
14. He perc_____ (感觉到, 看到) a light at the end of the tunnel.
15. Children are not independent. They are att_____ (黏着, 紧挨) to their parents.
16. She showed off her sle_____ (细长的, 苗条的) figure in a long narrow dress.
17. She has been changing partners often because she cannot have a sta_____ (稳定的) relationship with one person.
18. You must wear a bathing suit on a public beach. You're not allowed to bath na_____ (裸体的, 不穿衣服的).

Part C

1. Soldiers usually swear an oa_____ (誓言) of loyalty to their country.
2. The voter placed the ball_____ (投票) in the box.
3. They keep their valuables in a vau_____ (金库) at the bank.
4. A bird perched at the window led_____ (架, 突出物, 窗台).
5. The kitten is playing with a ball of ya_____ (纱, 线, 毛线).
6. The thieves have forced an ent_____ (入口) into the building.
7. The small hill was really a burial mou_____ (小丘, 沙堆, 墓冢).
8. We decided to celebrate New Year's E_____ (除夕夜) together.
9. The soldier was asked to choose between infantry (步兵队) and cav_____ (骑兵队).
10. This is a complex problem that is difficult to compr_____ (理解).
11. The angry crowd sho_____ (用力推, 推开, 推挤) the prisoner as he was leaving the court.
12. Don't pay attention to this rude remark. Just ig_____ (忽视, 不理) it.
13. The management held a secret meeting. The issues discussed were not disc_____ (揭露, 泄露, 显露) to the workers.
14. We could hear the sergeant (陆军中士) bel_____ (吼) commands (命令) to the troops.
15. The boss got angry with the secretary and it took a lot of tact (机智) to soo_____ (使.....震惊, 使.....缓和) him.
16. We do not have adeq_____ (足够的) information to make a decision.
17. She is not a child, but a mat_____ (成熟的) woman. She can make her own decisions.
18. The prisoner was put in soli_____ (独自的, 一人的) confinement (关押).

Part D

1. There has been a recent tr_____ (倾向, 趋势) among prosperous families

- toward a smaller number of children.
2. The ar_____ (区域, 面积) of his office is 25 square meters.
 3. Phil_____ (哲学) examines the meaning of life.
 4. According to the communist doc_____ (信条, 教义), workers should rule the world.
 5. Spending many years together deepened their inti_____ (亲密, 亲近).
 6. He usually read the sports sec_____ (部分) of the newspaper first.
 7. Because of the doctors' strike, the cli_____ (诊所) is closed today.
 8. There are several misprints on each page of this te_____ (课文).
 9. The suspect (嫌犯) had both opportunity and mot_____ (动机) to commit the murder.
 10. They insp_____ (检查, 视察) all products before sending them out to stores.
 11. A considerable amount of evidence was accum_____ (积累) during the investigation.
 12. The victim's shirt was satu_____ (浸透) with blood.
 13. He is irresponsible. You cannot re_____ (依靠) on him for help.
 14. It's impossible to eva_____ (评价) these results without knowing about the research methods that were used.
 15. He finally att_____ (达到, 获得) a position of power in the company.
 16. The story tells about a crime and subs_____ (随后的, 作为结果的) punishment.
 17. In a hom_____ (相同种类的, 同性质的) class all students are of a similar proficiency (水平).
 18. The urge to survive is inh_____ (天生的, 先天的) in all creatures.

Part E

1. The baby is wet. Her dia_____ (尿片) needs changing.
2. The prisoner was released on par_____ (假释).
3. Second year university students in the US are called soph_____ (大二学生).
4. Her favourite flowers were or_____ (兰花).
5. The insect causes damage to plants by its toxic (有毒的) sec_____ (分泌物).
6. The evacu_____ (撤离, 疏散) of the building saved many lives.
7. For many people, wealth is a prospect of unimaginable felic_____ (幸福).
8. She found herself in a pred_____ (困境) without any hope for a solution.
9. The deac_____ (执事, 信徒代表) helped with the care of the poor of the parish (教区居民).
10. The hurricane whi_____ (鞭打, 席卷) along the coast.
11. Some coal was still smol_____ (闷烧) among the ashes.
12. The dead bodies were mutil_____ (把.....切断) beyond recognition.
13. She was sitting on a balcony and bas_____ (晒太阳) in the sun.
14. For years waves of invaders pill_____ (掠夺) towns along the coast.
15. The rescue (营救) attempt could not proceed quickly. It was imp_____ (妨碍) by bad weather.

16. I wouldn't hire him. He is unmotivated and indo_____ (懒惰的) .
17. Computers have made typewriters old-fashioned and obs_____ (赶不上时代的) .
18. Watch out for his wil_____ (诡计多端的，狡猾的) tricks.

Appendix C

The Lex30 Test

请写出你从该词所能联想到的任何词（任何词都可，没有限制）。请不要翻查字典。

例如：feel sense numb smooth peel

attack	_____	_____	_____	_____
board	_____	_____	_____	_____
close	_____	_____	_____	_____
cloth	_____	_____	_____	_____
dig	_____	_____	_____	_____
dirty	_____	_____	_____	_____
disease	_____	_____	_____	_____
experience	_____	_____	_____	_____
fruit	_____	_____	_____	_____
furniture	_____	_____	_____	_____
habit	_____	_____	_____	_____
hold	_____	_____	_____	_____
hope	_____	_____	_____	_____
kick	_____	_____	_____	_____
map	_____	_____	_____	_____
obey	_____	_____	_____	_____
pot	_____	_____	_____	_____
potato	_____	_____	_____	_____
real	_____	_____	_____	_____
rest	_____	_____	_____	_____
rice	_____	_____	_____	_____
science	_____	_____	_____	_____
seat	_____	_____	_____	_____

spell _____
substance _____
stupid _____
television _____
tooth _____
trade _____
window _____

Appendix D

Writing Task One

Please write a composition of no less than 200 words according to the topic below. 请不要翻查字典。

经济发展往往是衡量一个国家成功与否的标准之一。但是也有人认为其它因素也很重要。你认为还有哪些因素？哪个因素是最重要的？

Appendix E

Writing Task Two

Please write a composition of no less than 200 words according to the topic below. 请不要翻查字典。

好的性格是帮助人们取得成功的因素之一。但是一些人认为其他因素也很重要。你认为还有哪些因素？哪个因素是最重要的？

Appendix F

Word Association Test One

一、请写出你从该词所能联想到的任何词（没有任何限制，任何词均可，请快速地写下自己的第一反应，在三分钟左右完成。请不要翻查字典。）

educate (v.) _____

afraid (adj.) _____

damage (v.) _____

quickly (adv.) _____

especial (adj.) _____

shout (n.) _____

happy (adj.) _____

二、下列词符合哪一个选项，请选出（如果选的是III., IV., 或V., 请在后面的横线上按该选项的要求填空）。不要翻查字典。

I I don't remember having seen this word before.

II. I have seen this word before but I don't know what it means.

III. I have seen this word before and I think it means _____ (近义词或翻译).

IV. I know this word. It means _____ (近义词或翻译).

V. I can use this word in a sentence. Eg: _____ (请造句).

例如：apple V. This apple tastes sweet and sour.

educate (v.) _____

afraid (adj.) _____

damage (v.) _____

quickly (adv.) _____

especial (adj.) _____

shout (n.) _____

happy (adj.) _____

Appendix G

Word Association Test Two

一、请写出你从该词所能联想到的任何词（没有任何限制，任何词均可，请快速地写下自己的第一反应，在三分钟左右完成。请不要翻查字典。）

civilize (v.) _____

cowardly (adj.) _____

wreck (v.) _____

amused (adj.) _____

hastily (adv.) _____

peculiar (adj.) _____

roar (n.) _____

二、下列词符合哪一个选项，请选出（如果选的是III., IV., 或V., 请在后面的横线上按该选项的要求填空）。不要翻查字典。

I I don't remember having seen this word before.

II. I have seen this word before but I don't know what it means.

III. I have seen this word before and I think it means _____ (近义词或翻译).

IV. I know this word. It means _____ (近义词或翻译).

V. I can use this word in a sentence. Eg: _____ (请造句).

例如：apple V. This apple tastes sweet and sour.

civilize (v.) _____

cowardly(adj.) _____

wreck (v.) _____

amused (adj.) _____

hastily (adv.) _____

peculiar (adj.) _____

roar (n.) _____